Karl Ulrich Mayer, Nancy Brandon Tuma (Eds.)

APPLICATIONS
OF EVENT HISTORY ANALYSIS
IN LIFE COURSE RESEARCH

Contents

III Methodological issues

Foreword

The use of multivariate stochastic models of change in discrete variables in continuous time—in short, event history analysis—has been one of the major innovations in longitudinal analysis in recent years. It has affected demography, economics and sociology alike. In fact, it has effectively broken down century-old fences between these disciplines.

Initially, a major effort was made to develop and explicate the methodology of event history analysis (see Tuma and Hannan, 1984; Heckman and Singer, 1985; Blossfeld, Hamerle, and Mayer, 1986). Paramount concerns at the early stage were to sort out several types of time-dependence, to struggle with the problem of population heterogeneity, to bring substantive theory and the statistical models into agreement, and to solve problems of estimation. Not the least of social scientists' problems in this area were such stumbling blocks as collecting or finding relevant data with continuous event histories, organizing and retrieving nonrectangular data files, making sense out of strange coefficients, and developing and evaluating software for these new forms of analysis.

The time came to go one step further and to make a sober assessment of the contribution of event history analysis to substantive problems. To this end, we convened a conference on applications of event history analysis in research on the life course. The conference took place at the Max Planck Institute for Human Development and Education in Berlin on June 5-7, 1986. The context of the conference was fourfold:

1) At the Max Planck Institute for Human Development and Education, there is currently a research program on "Life Course and Social Change" within which retrospective life history data on five German birth cohorts (those born in 1929-31, 1939-41, 1949-51 [already collected], 1919-21 [in the field], and 1959-61 [planned for 1988-89]) are being collected and analyzed. Those involved in this research program have a special stake in using the best methods for analyzing these data to investigate the processes of job and class mobility, family formation and fertility, migration, women's employment, and so forth.

2) The U.S. Social Science Research Council's Committee on Comparative Stratification through its sub-committee on life history research (chaired by David L. Featherman, University of Wisconsin at Madison) has, for some time, fostered the use of event history analysis for cross-national comparisons. As a result, the Social Science Research Council made a portion of a grant from the U.S. National Science Foundation available to support this conference by funding the attendance of young scholars.

3) A few years ago, the Frankfurt-Mannheim-Berlin based Special Research Unit on Microanalytic Founda-tions of Social Policy (Sonderforschungsbereich 3) initiated a large-scale socio-economic panel involving around 6,000 households and 12,000 individuals in West Germany. The prospect of analyzing these new data has led this group to make a major investment in the methodology of longitudinal analysis.

4) Finally, the conference was one of the annual meetings of the informal German working group on mathematical sociology (MASO), which has a long-term commitment to fostering formal modeling in substantive sociological research.

These networks produced an unusually rich and intense conference. It included all of West Germany's event history analysts and a strong group of its international protagonists. The conference clearly provided a sense of progress in the area. Social mobility research, for instance, is beginning to give up the conventional static comparisons of statuses held for very unequal durations. Increasingly time-dependent explanatory variables are being introduced—leading to results that are often not yet fully understood. "Piece-wise constant" models of hazard (or transition) rates were surely the "winners" of the day from a modeling perspective. During this meeting, one of the first analyses in which job trajectories were investigated in the context of time-dependent economic and societal quarter-year time series that describe historical conditions was also presented.

This volume in the series of "Materialien aus der Bildungsforschung" comprises all of the papers presented at the conference and a selection of the comments on the papers by the discussants. Many of the papers have been revised after the conference and, therefore, a few of the comments, while valid as arguments, do not apply well to the revised versions. We hope that these proceedings will contribute to a wide circulation of the contributions made at this conference.

I Job histories and occupational careers

Career opportunities in the Federal Republic of Germany. A dynamic approach to study life course, cohort, and period effects

Hans-Peter Blossfeld

1. Introduction

After a long period in which sociologists have concentrated their discussions on inter-generational mobility (Payne and Payne, 1983; Müller, 1985; Teckenberg, 1985), there is now a growing interest in studying and explaining individual career opportunities (see, e.g., Sørensen, 1975, 1977; Sørensen and Tuma, 1981; Silerman, 1977; Tuma, 1986; Carroll and Mayer, 1986; Mayer and Carroll, 1986). This recent research conceives careers as determined by two sets of variables: variables describing characteristics of individuals (e.g., family background, ability or education) and variables capturing attributes of the labor market structure (e.g., occupational groups, social class, organizational size or industry classification). There are many empirical studies that try to separate the relative importance of these individual and structural factors, but they tend to treat the structural side of the problem as more or less time-constant.

Certainly, any idea of social structure includes the notion of relative

stability (Weber, 1976; Parsons, 1951, 1959; Giddens, 1973), but this does not mean that the structure stays completely unchanged over time (Lockwood, 1956; Guessous, 1967). In the case of the labor market, structural changes seem to be the rule rather than the exception. All modern societies show a rapidly changing occupational distribution and a shift in employment from primary to secondary sector, and from secondary to tertiary sector (Müller, 1983; Haller, 1982, 1983, 1986). Educational expansion leads to labor market imbalances at all levels of qualification (Blossfeld, 1983, 1984a, 1984b, 1984c; Müller and Mayer, 1976). In general, both the nature of positions and the order in which they are occupied are changing (Blossfeld, 1985a, 1985b; Mayer 1977a, 1977b, 1978, 1979). Thus, a comprehensive explanation of how career opportunities are generated and how people are matched to jobs has to take the changing labor market structure into account.

There are mainly two ways in which a changing labor market structure affects career opportunities. The first is that people start their careers in different structural contexts. It has been often supposed that these specific historic conditions at the point of entry into the labor market have a substantial impact upon the people's subsequent careers (Hogan, 1981; Ornstein, 1976; Blossfeld, 1986b, 1986c, 1986d; Müller, 1978). This kind of influence is generally called a cohort effect (Ryder, 1965; Mason et al., 1973; Carlsson and Karlsson, 1970; Glenn, 1977). The second way changing labor market structure influences career opportunities is that it improves or worsens the career prospects of all people within the labor market at a given time. For example, in a favorable economic situation with low unemployment, there will be a great number of opportunities. This kind of influence is generally called a period effect.

While both cohort and period effects are structural sources of over-time variation in attainments, life-course related variations in socio-economic outcomes are mostly regarded to reflect changes in the individual (e.g., changes

in skills, knowledge, or experience) (Mincer, 1974) or as a result of the oppor-
tunity structure of a given society (Sørensen, 1984). In any case, if cohort,
period, and life course effects have distinct interpretations, then an analysis
which omits one of these dimensions will quite likely produce spurious find-
ings.

The purpose of this paper is, therefore, to study the causal impact of
cohort, period, and life course effects on career opportunities of German males.
I am especially interested in showing how different cohorts enter the labor
market under different structural contexts and how these specific entries af-
fect the subsequent careers of individuals under changing labor market condi-
tions. Using a large life history dataset on the full career trajectories of
German respondents from the birth cohorts 1929-31, 1939-41 and 1949-51 up to
age at interview together with several time series indicating the development
of the social and economic structure of the Federal Republic of Germany, I try
to separate these dimensions. I first discuss several labor market theories
and suggest hypotheses about the influence of cohort, period, and life course
effects. I then describe the data, variables and models I use to examine the
hypotheses. Finally, I report the results of the analyses.

2. Hypotheses About Time-Dependence of Career Opportunities

Why some people have better jobs than others has long been a key issue in
the work of sociologists and labor market economists. In the sixties and early
seventies, status attainment theory (see, e.g., Blau and Duncan, 1967; Sewell
and Hauser, 1975; Müller, 1972, 1975) drew attention to education, the father's
position, and other background variables. Introducing path analysis with linear
regression, this research provided quantitative estimates of the relative im-
portance of education and social origin, while largely ignoring any temporal
nature of the attainment process. As Sørensen expressed it: The effects of

variables are established ". . . whether the current occupation of the respondents is the first or the one they are holding after fifty years in the labor market" (Sørensen, 1985:24). Consequently, status attainment theory predicts a time-constant rate of leaving a job, depending only on education, social origin and other background variables. Furthermore, it assumes that positions in the labor market are freely available to anyone with the necessary resources and that changes in the structure of the labor market are not important. I shall refer to this theory as a static approach (Table 1).

Insert Table 1 about here

Another supply-side oriented model to explain the attainment process is human capital theory (Becker, 1975; Mincer, 1974). It postulates that levels of attainment reflect differences in productivity, and productivity is a result of investment in human capital. Hence, general education as well as ability leads to gains in attainment. While both are regarded as relatively constant over the attainment process, the main source of change in career opportunities is generated by a change in general on-the-job training and labor force experience. But, training after entry into the labor market is not evenly distributed over the life course (Mincer, 1974). Human capital theory predicts that people invest in their resources as long as their expected returns exceed their expected costs. Therefore, training is concentrated mainly in earlier phases of employment, where more time if left to recover training costs. In this way, the attainment process is time-dependent, since time in the labor force reduces the likelihood of new training and gains of attainment.

These predictions of human capital theory are based on the assumption that the labor market is perfectly competitive. But this is rarely the case. Tuma (1986) has shown that under the absence of perfect information and with the

costs accruing to searches, human capital theory leads to specific hypotheses
about the effects of rewards and resources on upward and downward moves.

If there is imperfect information and searches are not free, employers
must use observable indicators of a person's productivity to select employees,
and laborers must use observable indicators of job's rewards to select jobs.
In this situation, job-person mismatches may occur and job shifts are used to
resolve discovered mismatches. Because unexpected gains of these mismatches on
average balance unexpected losses, upward shifts and downward moves should be
more or less the same. But downward shifts should be more likely if there are
unexpected rewards, and unexpected rewards are the higher observable job re-
wards and the worse observable personal resources are. Conversely, upward
shifts are more likely if there are unexpected losses, and unexpected losses
are more likely the less observable job rewards and the better observable per-
sonal resources are.

In general, the occurrence of a job shift implies that either the employer
or the employee is not in equilibrium, and job shifts do not occur in equilib-
rium because no one can improve his present situation. In addition to imper-
fect information and search costs there may be other exogenous changes, for
example, economic or social shocks, new technologies, etc. But human capital
theorists do not explain why and how frequently these disequilibrating changes
occur. Hence, changing labor market structures and their effects on job mobil-
ity are outside the scope of their analysis. Because only time in the labor
force is regarded as a source of time-dependence of the attainment process,
I shall refer to human capital theory as a semi-static approach (Table 1).

Not everyone agrees that either the employer or the employee can start and
terminate the job-person match at any time. According to vacancy competition
theory (Sørensen, 1977), positions in the labor market are not freely available
to anyone with the necessary personal resources, because employees have control

over the decision to leave their job. This control over the decision to leave a
job is derived from several sources, for example, from job specific skills,
collective action, etc. Therefore, it is not always possible for the employer to
replace a current employee with another of higher productivity, even if such a
job candidate is available.

When employees have control over decisions to leave jobs, they will only
leave jobs when a better job is available. Hence, the vacancy competition
theory is a model about how gains in socio-economic attainments come about.
Downward moves are assumed to be the exception and lateral moves are regarded
as noise.

In general, positions will be available only when they are vacant. This
means that the creation of vacancies, not changes in personal resources, is the
central mechanism of mobility. A person may get access to a vacancy in a better
job without having increased his resources, and persons who increase their re-
sources may not be able to move to a better job, if there is no vacancy to
which to move.

Personal resources are only relevant as far as they serve to rank job
seekers in a labor queue. The criteria that serve to rank persons are indica-
tors of the person's trainability that are more or less fixed over a person's
career. Therefore, upward mobility results from an interaction between the
emergence of vacancies and individual differences in the ability to take ad-
vantage of these opportunities. For a given level of resources, opportunities
for even better jobs decline as the level of attainment already achieved in-
creases. It reflects a decreasing gap between actual and expected rewards, and
not an increase in personal resources. That means, that the likelihood to move
upward will decline with time in the labor force.

While the theory of vacancy competition brings the structure of positions
into the analysis of social mobility, it does not explain how structural changes

of the labor affect opportunity chances in terms of cohort and period effects. Vacancy theory only assumes that expansion and contraction of the system produces a constant rate of vacancies over all hierarchical levels. People entering the labor market are supposed to be distributed randomly among positions according to their level of resources, and first jobs are always lower than expected. Hence, vacancy competition theory is a model of upward mobility within a given structure of inequality and, therefore, also only a <u>semi-static approach</u> (Table 1). What is needed is a model that also takes into account the changing labor market structure.

The notion of structural change is by no means new in sociological mobility research. In analyzing square tables, cross-classifying the social class or occupational position of fathers and sons, there is a long history of attempts to separate the amount of "exchange mobility" from the mobility due to changes in social structure (cf., e.g. the early studies of Rogoff [1953] or Glass [1954], and the more recent works of Hauser [1977] or Erikson and Goldthorpe [1985]). But these attempts have not been very convincing, as they give a misleading impression of the role that structural changes play in career mobility.

The standard mobility table is obtained from a cross-sectional sample of men who are asked about their current position and about their fathers position when they were growing up. The father's position is normally referred to as the origin position, and the son's as the destination position. It is often seen as providing a picture of the social structure in two points in time. But this is a misleading picture, as already shown by Duncan (1966) and Sørensen (1985).

As fathers have sons at different ages and at different stages in their careers, the marginal distribution of origin positions does not correspond to a real distribution of origins in social structure at any point in time. And, as sons are in different stages of their careers at time of interview, the margi-

nal distribution of destinations does not correspond to a real distribution of destinations in social structure. "The result is that the typical mobility table aggregates career mobility processes for two generations spanning a large part of the century" (Sørensen, 1985:16). It simply ignores the time-dependent nature of the attainment process. The attainment process is not only dependent on the length of time in the labor force (Mincer, 1974; Sørensen, 1983), but also on specific historic conditions at time of entry into the labor market as well as on the specific labor market structure in every point of historical time.

Today, explanations of the direction of structural changes of the labor market and the development of job requirements are controversial among sociologists. Derived from case studies, industrial sociologists have postulated not only a general trend to occupational upgrading (Blauner, 1975), but also a trend to occupational downgrading (Bright, 1966), to a polarization of job requirements (Kern and Schumann, 1970, 1984), and to a diversification of skills in the course of development of technologies and work organizations. On a macrosociological level there is also the theory of Bell (1975) that claims an upgrading of employment structures, and the theory of Braverman (1977) that predicts a degradation of labor in the course of the development of modern societies.

Despite these different theoretical viewpoints, the trends displayed in the official employment statistics of the more advanced industrial nations show with an impressing regularity ". . . that the greatest increases in nonmanual employment over recent decades have occurred not in relatively low level clerical, sales and personal service grades but rather in professional, administrative and managerial occupations; and further, that the major decreases in manual employment have been in the less skilled rather than the more skilled categories" (Goldthorpe, 1986:10). In this sense, one can speak of modernization not only in terms of modern ideas, new life styles and changing tastes, but

also in terms of the labor market insofar as it is connected to a changing

occupational distibution and to a shift in employment from primary to secondary,

and from secondary to tertiary sectors (Rostow, 1966; Fourastie, 1954; Flora

1974, 1975; Zapf, 1983). Modernization means higher productivity and, therefore,

more white-collar employees and civil servants and less manual workers (Bell,

1975). On average, modernization of the labor market leads to an improvement

of the occupational structure, and hence to more opportunities (Smelser and

Lipset, 1966).

On the other side, in most Western societies we observe also a large-scale

and long-term unemployment as well as a growth of 'non-standard' employment

forms, for example, part-time work, temporary work, home-work etc. (Goldthorpe,

1986). These are connected mainly with the slowdown in economic activity and

contradict the thesis that at stage of advanced industrialism upgrading pro-

ceeds uniformly in all aspects of employment alike (Goldthorpe, 1986:14).

Therefore, a dynamic approach (Table 1) to the attainment process should bring

in both the increasing path of modernization and the cyclical development of

labor market conditions around this trend. Both aspects of the structural

change have an impact on entry into the labor market and on intra-generational

mobility.

Let me first consider possible effects of a changing labor market struc-

ture on entry into the labor market. People start their careers in different

structural contexts. The main feature of these contexts can be described by the

historic level of modernization and by the historic labor market conditions.

With Janossy (1966) we assume that modernization is connected with new job

requirements, and that these job requirements differ from the labor force mem-

ber's structure of qualification. Consequently, there is a continuously greater

difference between the structure of qualification available and the structure

of requirements needed by positions. An adaptation of both structures may be

reached only if labor market members acquire new skills or new entrants get access to these new jobs. With increasing age, the possibility of retraining is limited, because of decreasing returns of education. Therefore, labor market entrants are most likely to get access to these vacancies created by the process of modernization. As the level of modernization increases, cohorts of entrants should enter the labor market on higher average levels of attainment. But, the higher the attainment level at entry into the labor market, the less likely it will be to find a better job. Hence, we expect a decreasing rate of upward mobility with increasing levels of modernization at entry into the labor market (Table 1).

As these new entrance-positions are "modern position" in the sense of currently needed occupations, it is to be expected that incumbents are more protected from dismissals (because of rationalization) the higher the level of modernization at entry into the labor market. Therefore, we expect decreasing rate of downward moves with increasing entry-levels of modernization (Table 1).

If we look on the effects of the labor market conditions at time of entry into the labor market, it is reasonable to assume that the better the employment situation, the easier it is for entrants to reach higher levels of first jobs and the harder it will be to move up further. Therefore, the labor market conditions at entry into the labor market should be connected negatively with upward mobility (Table 1). If labor market entrants are dependent upon vacancies in jobs, their alternatives may be to wait until a suitable job becomes available or to accept a job whose attainments are less than expected. Given these two alternatives, most people will choose the latter. In this situation, unexpected rewards that lead to downward mobility are the less likely, the better the labor market conditions at time of entry are. Hence, we expect that the better the conditions of labor market at entry, the less downward moves will happen in the later career (Table 1).

Changing labor market structure not only has effects on entry into the labor market, but also affects the <u>actual career opportunities of all labor market members</u> and does them make also time-dependent.

If we consider modernization it is reasonable to assume that modernization leads on average to an improvement of the occupational structure, and to more opportunities (Smelser and Lipset, 1966) (Table 1). But it is not clear how this change comes about. For example, Blauner (1975) argues that modern technology and work organization always will lead to more skills and better employment. Whereas Kern and Schumann (1970) claim that modern technology and work organization lead to an occupational upgrading, producing higher and lower attainments at the same time. They call their theory polarization of job requirements. If Blauner is right, modernization should lead only to more upward moves; if Kern and Schumann are right, modernization leads to both, upward and downward moves. As some people can take advantage of higher attainments and some people are forced to move down. In the latter sense, modernization means simply more mobility with an average upgrading of the occupational structure (Table 1).

While modernization may lead to upward and downward mobility, it is relatively clear that better actual labor market conditions can only lead to more opportunities, and hence to more upward moves (Table 1).

Our discussion shows that a changing labor market structure should have strong cohort and period effects that create opportunities and force people to move in worse positions. This means that the creation of vacancies and the loss of positions in the course of structural changes count among the central mechanisms of career mobility, and affect people's actual mobility chances. In this light, the distinction of structural and exchange mobility in standard mobility tables, where only the latter is regarded as an indicator of the openness of society, seems to be obsolete. Due to structural changes, a person may get

access to a better position without having increased his resources, and persons who increase their resources may not be able to move to a better job, if there is no effect of social structure. It is clear that the process of attainment is time-dependent in a threefold sense: It depends on the historical time of entry, it depends on the time spent in the labor force and it depends on the actual historical time. Any comprehensive analysis of career opportunities should therefore also incorporate the influences of these effects.

Because of their linear dependency, an identification problem arises whenever one tries to estimate models with cohort, life course and period effects (Mason et al., 1973; Glenn 1977). There are at least two ways to solve the problem: "One can assume that one (or more) of these measures of time has no effect on the outcome, impose a sufficient number of equality constraints on the model to identify it, or include measures of the causal variables for which age, period, and cohort are surrogates" (Tuma and Hannan, 1984:192). I will follow the latter method and use time series to measure cohort and period effects in the empirical analysis.

3. Data and Methods

The hypotheses in Section 2 are tested using life histories collected between October 1981 and May 1983 by Karl Ulrich Mayer (1984a, 1984b). These life-history data include 2,171 German respondents from the cohorts 1929-31, 1939-41, and 1949-51. This sample is representative of the native-born German population of the Federal Republic of Germany (Brückner et al., 1984; Blossfeld, 1986a). The objective of this data collection effort was to record the life histories of the respondents over all sociologically relevant domains of life (social background, education, family, dwelling, etc.). The data were collected retrospectively by asking the respondents to reconstruct with exact dates their life histories within these domains of life. Recall errors are likely, espe-

cially if the events took place a long time ago. But there are few methodological cal studies about the reliability of retrospectively recorded data (Moss and Goldstein, 1979; Papastefanou 1980; Tölke, 1980). The general reservation with respect to the quality of such data is therefore plausible but not empirically demonstrated. It seems, however, that data on general schooling and occupational training events as well as the occupational history are relatively reliable (Tölke, 1980; Blossfeld, 1986a).

This data base constitutes a unique source of information about educational and occupational histories of men and women from selected birth cohorts and can be used to address the hypotheses set out in the theoretical section. Women were excluded as they show very different occupational structures and mobility patterns. The basic explanatory variables are job rewards, personal resources, and the changing labor market structure. Table 2 gives definitions and simple descriptive statistics for these variables.

The primary measures of life course effects is time in the labor force or labor force experience (Table 2). Human capital economists often use this variable to measure general on-the-job training (Mincer, 1974). Whereas in the vacancy competition model it measures the gap between actual and expected rewards (Sørensen, 1984).

Insert Table 2 about here

Education is used as a measure of observed personal resources. It combines the highest levels of general education and vocational traiing at entry into the job (Table 2).

As a measure of the attainment level or the job rewards occupational prestige is used, which is measured here by Wegener's (1985) score (Table 2). Compared to the wage rate, prestige reflects also non-pecuniary rewards of a job

and is therewith a better measure of the "goodness" of jobs (Goldthorpe and

Hope, 1972; Sørensen, 1977).

As a model using life course, cohort and period effects causes an iden-

tification problem, we use time series to indicate the development of the

labor market structure. But, as everyone knows, time series often measure

similar features of a process and are highly correlated. Therefore, another

identification problem invariably arises whenever these time series are in-

cluded in a model at the same time. One strategy to solve that problem is

to use only one of these series, or to choose only uncorrelated time series.

The problem is that time series chosen under this perspective may only cap-

ture specific parts of the development. If time series represent aspects

of an underlying regularity, it is more appropriate to search for components

that may be taken as source variables accounting for the observed develop-

ments. A statistical method to do that is the factor analysis (Weber, 1974).

We used 14 different time series to indicate social and economic structure's

development. Employing factor analysis with principal factoring and equimax

rotation gives us two orthogonal factors which explain 96.2 percent of the var-

iance of the 14 time series (see Table 3).

Insert Table 3 about here

There is a very high correlation between factor 1 and the level of produc-

tivity, the national income per capita, the national income per economically

active person, the private consumption, the proportion of expenditures for ser-

vices of the private consumption, the proportion of gainfully employed in the

public sector, the proportion of 13-year-old pupils attending "German Gym-

nasium," the proportion of gainfully employed in the service sector, the propor-

tion of students of the resident population, the proportion of civil servants

of the economically active population, and the proportion of white-collar em-
ployees of the economically active population. All these variables can be seen
as indicators of the modernization process in the German society; hence, I call
the first factor "level of modernization."

Factor 2 is highly correlated with the (negative) unemployment rate, the
proportion of the gross national product for investments in plants and equip-
ment, and the proportion of registered vacancies of all dependent jobs. All
three variables are clearly measures of the cyclical development of the labor
market; therefore I call it "labor market conditions."

As for the German case, reasonable time series only from 1950 to the time
of interview are available, we computed factor scores for each of these years.
To get also factor scores for the period between 1945 and 1950, we used a poly-
nomial regression to extrapolate the factor values. Factor scores for the
"level of modernization" were predicted with an R^2 of 0.996 and factor scores
of the factor "labor market conditions" were predicted with an R^2 of 0.793.
The results of the whole procedure are shown in Figures 1 and 2.

Insert Figures 1 and 2 about here

The "level of modernization" (Figure 1) shows a trend with a slightly in-
creasing slope, and the "labor market conditions" show a cyclical development,
with downturns around 1967 and 1973, both in accordance with historical events
(Figure 2). As both factors are orthogonal, it is possible to use both varia-
bles simultaneously within an equation to represent the development of the
labor market structure, and to estimate the cohort and the period effects. To
measure the cohort effects, I used the factor scores "level of modernization"
(Figure 1) and "labor market conditions" (Figure 2) of the year persons entered
the labor force. To introduce the period effects, I used the method of episode-

splitting (Peterson, 1986; Blossfeld, Hamerle and Mayer, 1986). Here for each interval of time within a job where the factor scores "level of modernization" and "labor market conditions" stayed constant a separate record of data was created. In each record four pieces of information are given: (1) the duration in the job at the beginning and end of the interval to which the record pertained; (2) whether the interval ended with a job mobility or not; (3) the values of the factor scores "level of modernization" and "labor market conditions" at the beginning of the interval; (4) the values of the other covariates relevant for the analysis. This leads to 22,843 records. In the computations each record was treated as a separate observation. It is important to note that this procedure for setting up the data does not change the times of job episodes and therefore does not affect the estimates of the other covariates.

For the directional movements, I defined prestige increases of 10 percent and greater as upward moves, prestige increases between 0 and 10 percent as lateral moves, and all prestige decreases as downward moves.

To study the career mobility, I use as the dependent variable the instantaneous rate of transition between jobs (Tuma and Hannan, 1984; Blossfeld, Hamerle and Mayer, 1986). My goal is to specify the rate as a function of personal characteristics, attainments of jobs, and the changing labor market structure. Here I use an exponential model to incorporate the time-dependent effects of the life course, the cohort, and the period:

$$r(t/\underline{x}(t)) = \exp(\underline{\beta}'\underline{x}(t))$$

where $\underline{x}(t)$ is the time-dependent vector of the exogenous variables.

4. Results

Let me begin the analysis with a description of the development of the average prestige scores of the birth cohorts 1929-31, 1939-41, and 1945-51 over the historical time (Figure 3). This shows that for each younger birth

cohort, the average prestige score starts on a higher level and has a greater slope. The cohort of 1939-41 surpasses cohort 1929-31 in 1967, and the cohort of 1949-51 surpasses cohort 1929-31 in 1977 and cohort 1939-41 in 1979. Compared with the development of modernization (Figure 3), entrance prestige score levels of the three birth cohorts lie around the path of modernization. This supports the hypothesis that after World War II each younger birth cohort entered the labor market on higher average attainment levels (Blossfeld, 1986c, 1986d). It seems that there are also life course effects, because each of these curves shows a typical rise. But, as Figure 3 presents only history-specific average prestige scores, age-specific patterns are hard to compare.

Another way to describe this development is therefore to show how the average prestige scores are distributed over age by birth cohort (Figure 4). Here, every cohort shows the typical increase in job rewards over a person's life cycle: the curves increase rapidly in the beginning and then gradually taper off. This confirms the thesis that there is a systematic direction in people's life course (Mincer, 1974; Sørensen, 1977). But again, every younger cohort reaches a specific average attainment level at an earlier age and shows different slopes, indicating that there are also cohort and period effects.

Insert Figures 3 and 4 about here

Therefore, a model is needed that separates the effects of the cohort, the period, and the life course. Furthermore, this model should be based on entrance cohorts and not on birth cohorts, because entrance into the labor market and not time of birth is the relevant event for the attainment process. As shown in Figure 5 there are ten and more years of variation in entrance into the labor market for the birth cohorts 1929-31, 1939-41, and 1949-51. Members of the same birth cohort will therefore start and experience their careers in different

structural contexts.

Insert Figure 5 about here

A model that tries to separate life course, cohort, and period effects as well as describes the cohorts in terms of differences in structural contexts at entry into the labor market is presented in Table 4.

Insert Table 4 about here

This table reports estimates of the rates for upward, lateral, and downward moves. The results for lateral moves are presented but not interpreted. "They evidently represent a mixture of gains and losses in job rewards other than those summarized by status" (Sørensen and Tuma, 1981:83).

All coefficients are metric coefficients. But I will not compare the relative magnitude of the effects of different variables within models. Instead, I rely on the significance level to indicate relative importance of variables. In evaluating a model's performance I also use a likelihood ratio test comparing the model to a baseline. The baseline for every directional move is model 1, the model of a constant rate. This test gives a chi-square value, reported in Table 4.

It is useful to demonstrate step by step that there are life course, cohort, and period effects. This procedure lets one begin with a simple model and introduce complexity gradually.

Model 2 has only one independent variable: time in the labor force at entry into the origin job i. For both upward and downward moves this model performs significantly better than the baseline model of a constant rate (model 1 in Table 4). The coefficient of time in the labor force is negative for upward

and downward moves. Based on our theoretical discussion in Section 2, there are only substantive interpretations of the effect of time in the labor force in upward moves (Table 1). According to human capital theory, training is concentrated mainly in earlier phases of employment where more time is left to recover training costs. Time in the labor force reduces the likelihood of new training and therefore for further gains in attainment. Another explanation is given by vacancy competition theory. Here, time in the labor force reflects a decreasing gap between actual and expected rewards, and not an increase in personal resources. For a given level of resources it gets harder to find an even better job.

Model 3 introduces education and prestige of the origin job i. For both upward and downward moves this model performs significantly better than model 2 indicating that at least one of the introduced two variables has a significant effect. Looking at the coefficient it can be seen that both variables are statistically significant at the 0.001 level.

The positive effect of education in upward moves is in accordance with the status attainment theory, the human capital theory, and the vacancy competition theory (Table 1). But all three theories have different explanations. Status attainment theory says, increases in educational attainment lead always to better job opportunities and have a positive effect in upward moves. Under the condition of an imperfect labor market, human capital theory predicts that upward shifts are the more likely, if there are unexpected losses, and unexpected losses are more likely the better observable personal resources are (Tuma, 1986). Education will therefore have a positive effect in upward moves. Finally, vacancy competition theory explains this positive effect of education with the ability of better educated persons to get better places in the labor queue and therefore to take more advantage of opportunities.

The positive effect of prestige in upward moves can also be explained with

two theories (Table 1). For Sørensen, opportunities for even better jobs decline as the level of attainment already achieved increases. It should be noted that within the framework of the vacancy competition theory the coefficient of prestige is also a measure of the opportunity structure of a given society: the larger the absolute magnitude of this coefficient, the fewer opportunities for gains are available in the studied society. Another explanation is given of the human.capital theory assumed that there are costs accruing to searches and there is no perfect information. In this case, upwardward moves are more likely if there are unexpected losses, and unexpected losses are more likely the less observable job rewards are. That means, the higher the prestige score of the origin job, the less upward moves are to be expected.

Also in accordance with this modified form of human capital theory (Tuma, 1986) are the negative effect of education and the positive effect of prestige in downward moves: Downward moves should be more likely if there are unexpected rewards, and unexpected rewards are the higher observable job rewards and the worse observable personal resources are. Because unexpected gains of mismatches on average balance unexpected losses, according to human capital theory upward shifts and downward moves should be more or less the same. Table 4 shows that this is the case.

Even if vacancy competition theory is able to explain the estimated coefficients as shown above, there are also results that may contradict this theory. Most important, human capital theory regards downward moves as an exception. But this is not the case in Table 4. The number of downward shifts (475) is nearly as important as the number of upward shifts (590). Also the effect of time in the labor force should not be significant, if education and prestige are controlled for in upward moves. Otherwise this variable is no adequate proxy of the discrepancy between resources and current job rewards.

It should be noted that human capital theory is interesting for sociologi-

cal mobility research only insofar as an imperfect labor market is assumed. As

shown by Tuma (1986), only if absence of perfect information and costs accruing

to searches are supposed, human capital theory leads to more specific hypotheses

about mobility. Otherwise the labor market should be in equilibrium, and in

equilibrium job shifts do not occur because no one can improve his present

sitution.

But, as argued above, there is also another shortcoming of human capital

theory and vacancy competition theory: Both theories are only semi-static

approaches and do not take into account the changing labor market structure.

Model 4 is an equation that incorporates the changing labor market struc-

ture at entry into the labor market with two variables: the level of moderni-

zation and the labor market conditions. For upward and downward moves this

model performs significantly better, indicating that there are cohort effects.

But as Table 4 shows, only the variable level of modernization at entry into the

labor market is significant in both directional moves.

The effect of labor market conditions at entry into the labor market is

only significant in model 5, where the period effects are also controlled for.

This model leads to a huge increase of the chi-square statistics. That is to

say that period effects count to the most important influences in occupational

mobility. In model 5 the modernization at entry into the labor market has a

negative effect in upward moves. That means, the higher the level of moderni-

zation, the better the entrance level (see Figures 3 and 4), and the less

likely are further upward moves. The same is true for the negative effect of

labor market conditions at entry into the labor market, if upward moves are re-

garded.

The level of modernization at entry into the labor market has also a nega-

tive effect in downward moves, because entrants move in modern occupations if

the modernization level is high, and this protects them from dismissal because

of rationalization. Negative is also the effect of labor market conditions at entry into the labor market in downward moves. This is the case, because unexpected rewards are the less likely the better the labor market conditions at time of entry are.

The period effect of modernization is positive in upward and downward moves, supporting the theory of polarization of job requirements. Modernization means more mobility at all. And, finally, the period effect of labor market conditions is also positive. The better the labor market conditions, the more opportunities are created. The empirical analysis shows therefore that there are strong cohort, period, and life course effects. Hence, the process of attainment is time-dependent: It depends on the time spent in the labor force, it depends on the historical time of entry into the labor market, and it depends on the actual historical time.

5. Conclusion

In this research I have undertaken three main tasks: (1) I have examined the implications for occupational mobility patterns of several models (status attainment theory, human capital theory, and vacancy theory) about the way persons are matched to jobs. (2) I have proposed introducing the changing labor market structure into intra-generational mobility research in order to study the time-dependent nature of the attainment process adequately. (3) I have used event history data and time series to solve the identification problem in models estimating cohort, life course, and period effects.

The theoretical discussion showed that there are three main approaches to the study of the intra-generational attainment process. First, there are models that ignore any temporal nature of the attainment process, for example, the status attainment theory. I called these models static approaches. Second, there are models that only regard time in the labor force as a source of time-

dependence. Human capital theory and vacancy competition theory count among these semi-static approaches. Finally, there are models that incorporate life course effects and the changing labor market structure in terms of cohort and period effects. I called these models dynamic approaches.

But, because of their linear dependency, an identification problem arises whenever one tries to estimate models with cohort, life course, and period effects. To solve this problem I used surrogates to measure cohort and period. Employing factor analysis I got two orthogonal factors that approximately describe the changing labor market structure after the Second World War in West Germany. The first factor was named "level of modernization" and the second one "labor market conditions." Using the factor scores of these factors I incorporated the changing labor market structure in a model estimating the instantaneous rate of transition between jobs.

Our empirical analysis provides evidence that there are strong life course, cohort, and period effects. The attainment process is time-dependent in a threefold sense: It depends on the time spent in the labor force, it depends on the historical time of entry into the labor market, and it depends on the actual historical time. Analyses of standard mobility tables that distinguish on the basis of cross-sections between structural mobility and exchange mobility give a misleading picture of the mechanisms of attainment. The creation of vacancies and the loss of positions in the course of structural change count to the central mechanism of career mobility, and affect people's actual mobility chances.

References

Becker G S. (1975): Human Capital, New York and London: Columbia University

Press.

Bell D. (1975): Die nachindustrielle Gesellschaft, Frankfurt: Campus Verlag.

Blau P M & Duncan O D. (1967): The American Occupational Structure, New York:

Wiley.

Blauner R. (1975): 'Entfremdung, Technologie und Arbeitsqualifikation bei

tendenzieller Höherqualifizierung der Gesamtarbeitskraft', in Hegelheimer

A (ed): Texte zur Bildungsökonomie, Frankfurt and Berlin: Verlag

Ullstein: 446-451.

Blossfeld H-P. (1983): 'Höherqualifizierung und Verdrängung - Konsequenzen der

Bildungsexpansion in den Siebziger Jahren', in Haller M & Müller W (eds):

Beschäftigungssytem im gesellschaftlichen Wandel, Frankfurt: Campus

Verlag: 184-240.

-- (1984a): 'Bildungsreform und Beschäftigung der jungen Generation im öffent-

lichen und privaten Sektor', Soziale Welt, 35: 159-189.

-- (1984b): 'Die Entwicklung der qualifikationsspezifischen Verdienstrelatio-

nen von Berufsanfängern zwischen 1970 und 1982', Kölner Zeitschrift für

Soziologie und Sozialpsychologie, 36: 293-322.

-- (1984c): 'Bildungsexpansion und Tertiarisierungsprozeß. Eine Analyse der

Entwicklung geschlechtsspezifischer Arbeitsmarktchancen von Berufsanfän-

gern unter Verwendung eines log-linearen Pfadmodells', Zeitschrift für

Soziologie, 13: 20-44.

-- (1985a): 'Berufseintritt und Berufsverlauf. Eine Kohortenanalyse über die

Bedeutung des ersten Berufs in der Erwerbsbiographie', Mitteilungen aus

der Arbeitsmarkt- und Berufsforschung, 18: 177-197.

-- (1985b): <u>Bildungsexpansion und Berufschancen</u>, Frankfurt and New York: Campus

    Verlag.

-- (1986a): 'Zur Repräsentativität der Sfb-3-Lebensverlaufsstudie. Ein Vergleich

    mit Daten aus der amtlichen Statistik', Frankfurt and Mannheim (forthcoming:

    <u>Allgemeines Statistisches Archiv</u>).

-- (1986b): 'New Methods to Study the Relationship Between Education and Occupa-

    tion', in Parkes L, Sellin B & Tessaring M (eds): <u>Education/Training and</u>

    <u>Labor Market Policy</u>, Gravenhage, The Netherlands: SVO: III-3-III-22.

-- (1986c): 'Labor Market Entry and the Sexual Segregation of Careers in the

    Federal Republic of Germany' (forthcoming: <u>American Journal of Sociology</u>).

-- (1986d): 'Entry into the Labor Market and Occupational Career in Germany in

    Comparison to American Studies' (forthcoming: <u>International Journal of</u>

    <u>Sociology</u>).

Blossfeld H-P, Hamerle A & Mayer K U. (1986): <u>Ereignisanalyse. Statistische</u>

    <u>Theorie und Anwendung in den Wirtschafts- und Sozialwissenschaften</u>,

    Frankfurt and New York: Campus Verlag.

Braverman H. (1977): <u>Die Arbeit im modernen Produktionsprozeß</u>, Frankfurt:

    Campus Verlag.

Bright J R. (1966): 'The Relationship of Increasing Automation and Skill Re-

    quirements', in The Employment Impact of Technological Change. Appendix

    vol 2: Technology and the American Economy, report of the National

    Commission on Technolgy, Automation, and Economic Progress, Washington,

    D.C.: 203-221.

Brückner E et al. (1984): <u>Methodenbericht "Lebensverläufe"</u>, Mannheim: ZUMA.

Carlsson G & Karlsson K. (1970): 'Age, Cohorts and the Generation of Genera-

    tions', <u>American Sociological Review</u>, 35: 710-718.

Carroll G R & Mayer K U. (1986): 'Job-Shift Patterns in the Federal Republic of Germany: The Effects of Social Class, Industrial Sector, and Organizational Size' (forthcoming: American Sociological Review).

Duncan O D. (1966): 'Methodological Issues in the Analysis of Social Mobility', in Smelser N J & Lipset S M (eds): Social Structure and Mobility in Economic Development, Chicago, Ill.: Aldine: 51-63.

Erikson R & Goldthorpe J H. (1985): Communality and Variation in Social Fluidity in Industrial Nations. Some Preliminary Results, CASMIN-Working Paper No. 4, Universität Mannheim.

Flora P. (1974): Modernisierungsforschung, Opladen: Westdeutscher Verlag.

-- (1975): Indikatoren der Modernisierung, Opladen: Westdeutscher Verlag.

Fourastie J. (1954): Die große Hoffnung des zwanzigsten Jahrhunderts, Köln: Deutz (frz. 1949).

Giddens A. (1973): The Class Structure of the Advanced Societies, London: Hutchinson.

Glass D V. (ed) (1954): Social Mobility in Britain. London: Routledge & Kegan Paul.

Glenn N D. (1977): Cohort Analysis, Beverly Hills, Cal.: Sage Publications.

Goldthorpe J H. (1986): Employment, Class and Mobility: A Critique of Liberal and Marxist Theories of Long-Term Change, paper for conference on Social Change and Development, Berkeley, August 26-28 1986, Oxford: Nuffield College.

Goldthorpe J H & Hope K. (1972): 'Occupational Grading and Occupational Prestige', in Hope K (ed): The Analysis of Social Mobility: Methods and Approaches, Oxford: Clarendon Press: 19-80.

Guessous M. (1967): 'A General Critique of Equilibrium Theory', in Moore W E & Cook R M (eds): Readings on Social Change, Englewood Cliffs, N.J.: Prentice-Hall: 23-35.

Haller M. (1982): 'Auf dem Weg zur Dienstleistungsgesellschaft?', Wirtschaft

　und Gesellschaft, 8: 607-654.

-- (1983): Klassenstrukturen und Mobilität in fortgeschrittenen kapitalistischen

　Gesellschaften, Habilitationsschrift, Universität Mannheim.

-- (1986): 'Sozialstruktur und Schichtungshierarchie im Wohlfahrtsstaat. Zur

　Aktualität des vertikalen Paradigmas der Ungleichheitsforschung', Zeit-

　schrift für Soziologie, 15: 167-187.

Hauser R M. (1977): 'A Structural Model of the Mobility Table', Social Forces,

　56: 919-953.

Hogan D P. (1981): Transitions and Social Change, New York: Academic Press.

Janossy F. (1966): Das Ende der Wirtschaftswunder, Frankfurt: Verlag Neue

　Kritik.

Kern H & Schumann M. (1970): Industriearbeit und Arbeiterbewußtsein, 2 vols,

　Frankfurt and Köln: Suhrkamp.

-- (1984): 'Neue Produktionskonzepte haben Chancen', Soziale Welt, 35: 146-

　158.

Lockwood D. (1956): 'Some Remarks on "The Social System"', The British Journal

　of Sociology, 7: 134-146.

Mason K O et al. (1973): 'Some Methodological Issues in Cohort Analysis of

　Archival Data', American Sociological Review, 38: 242-258.

Mayer K U. (1977a): 'Ungleiche Chancen und Klassenbildung', Soziale Welt, 28:

　466-493.

-- (1977b): Recent Developments in the Opportunity Structure of (West-)German

　Society 1935-1971, SPES-Working Paper No. 67, Universität Frankfurt and

　Universität Mannheim.

-- (1978): 'Wandel im Beschäftigungssystem, soziale Mobilität und berufliche

　Flexibilität', BeitrAB, 30: 235-267.

-- (1979): 'Strukturwandel im Beschäftigungssystem und berufliche Mobilität
zwischen Generationen', Zeitschrift für Bevölkerungswissenschaft, 5:
267-297.

-- (1984a): Lebensverläufe und Wohlfahrtsentwicklung. Bericht über die
Forschungstätigkeit in der zweiten Forschungsphase 1982-1984, Frankfurt
and Mannheim: 119-142.

-- (1984b): Lebensverläufe und Wohlfahrtsentwicklung. Antrag auf Förderung für
die dritte Forschungsphase 1985-1987, Frankfurt and Mannheim: 131-171.

Mayer K U & Carroll G R. (1986): Jobs and Classes: Structural Constraints on
Career Mobility, paper presented to the American Sociological Association,
August 26 to September 2, New York City, Berlin: Max Planck Institute for
Human Development and Education.

Mincer J. (1974): Schooling, Experience, and Earnings, New York: Columbia Uni-
versity Press.

Moss L & Goldstein H (eds). (1979): The Recall Method in Social Surveys.
London: University of London.

Müller W. (1972): 'Bildung und Mobilitätsprozeß - eine Anwendung der Pfad-
analyse', Zeitschrift für Soziologie, 1: 65-84.

-- (1975): Familie, Schule und Beruf. Soziale Mobilität und Prozesse der Status-
zuweisung in der Bundesrepublik, Opladen: Westdeutscher Verlag.

-- (1978): Klassenlage und Lebenslauf, Habilitationsschrift, Universität
Mannheim.

-- (1983): 'Wege und Grenzen der Tertiarisierung: Wandel der Berufsstruktur in
der Bundesrepublik Deutschland 1950-1980', in Matthes J (ed): Krise der
Arbeitsgesellschaft?, Frankfurt: Campus Verlag: 142-160.

-- (1985): 'Mobilitätsforschung und Arbeitsmarkttheorie', in Knepel H & Hujer R
(eds): Mobilitätsprozesse auf dem Arbeitsmarkt, Frankfurt: Campus Verlag:
17-40.

Müller W & Mayer K U. (1976): Chancengleichheit durch Bildung? Untersuchungen
über den Zusammenhang von Ausbildungsabschlüssen und Berufsstatus, Stutt-
gart: Klett Verlag (Deutscher Bildungsrat: Studien und Gutachten der
Bildungskommission, vol 42).

Ornstein M D. (1976): Entry into the American Labor Force. New York: Academic
Press.

Papastefanou G. (1980): Zur Güte von retrospektiven Daten - Eine Anwendung
gedächtnispsychologischer Theorie und Ergebnisse einer Nachbefragung,
Working Paper No. 29 of the Sfb 3 "Mikroanalytische Grundlagen der Gesell-
schaftspolitik", Frankfurt and Mannheim.

Parsons T. (1951): The Social System. Glencoe, Ill.: The Free Press.

-- (1959): 'General Theory in Sociology', in Merton R K et al (eds): Sociology
Today, New York: Basic Books: 3-38.

Payne G & Payne J. (1983): 'Occupational and Industrial Transition in Social
Mobility', The British Journal of Sociology, 34: 72-92.

Peterson T. (1986): Estimating Fully-Parametric Hazard Rate Models with Time-
Dependent Covariates by the Method of Maximum Likelihood (forthcoming:
Sociological Methods and Research).

Rogoff N. (1953): Recent Trends in Occupational Mobility, New York: The Free
Press.

Rostow W W. (1966): The Stages of Economic Growth, Cambridge: Cambridge Univer-
sity Press.

Ryder N B. (1965): 'The Cohort as a Concept in the Study of Social Change',
American Sociological Review, 30: 843-861.

Sewell W H & Hauser R M. (1975): Education, Occupation and Earnings, New York:
Academic Press.

Smelser N J & Lipset S M (eds). (1966): Social Structure and Mobility in Eco-
nomic Development, Chicago, Ill.: Aldine.

Sørensen A B. (1975): 'The Structure of Intragenerational Mobility', American Sociological Review, 40: 456-471.

-- (1977): 'The Structure of Inequality and the Process of Attainment', American Sociological Review, 42: 965-978.

-- (1983): 'Processes of Allocation to Open and Closed Positions in Social Structure', Zeitschrift für Soziologie, 12: 203-224.

-- (1984): 'Social Structure and Mechanisms of Life Course Processes', in Sørensen A B, Weinert F & Sherrod L (eds): Human Development: Multidisciplinary Perspectives (forthcoming).

-- (1985): Theory and Methodology in Stratification Research, Cambridge, Mass: Harvard University Press.

Sørensen A B & Tuma N B. (1981): 'Labor Market Structures and Job Mobility', Research in Social Stratification and Mobility, 1: 67-94.

Spilerman S. (1977): 'Careers, Labor Market Structure, and Socioeconomic Achievement', American Journal of Sociology, 83: 551-593.

Teckenberg W. (1985): 'Die Erwerbsstrukturkonzeption in Humankapital- und Statuszuweisungsmodellen. Einige theoretische Alternativen', Kölner Zeitschrift für Soziologie und Sozialpsychologie, 37: 431-460.

Tölke A (1980): Zuverlässigkeit retrospektiver Verlaufsdaten - Qualitative Ergebnisse einer Nachbefragung, Working Paper No. 30 of the Sfb 3 "Mikroanalytische Grundlagen der Gesellschaftspolitik", Frankfurt and Mannheim.

Tuma N B. (1986): Effects of Labor Market Structure on Job-Shift Patterns, Working Paper No. 83-11 of the International Institute for Applied Systems Analysis, Laxenburg (forthcoming: Heckman J J & Singer B (eds): Longitudinal Analysis of Labor Market Data).

Tuma N B & Hannan M T. (1984): Social Dynamics: Models and Methods, New York.

Weber E. (1974): <u>Einführung in die Faktorenanalyse</u>. Jena: VEBGustav Fischer

    Verlag.

Weber M. (1976): <u>Wirtschaft und Gesellschaft</u>, Tübingen: Mohr.

Wegener B. (1985): 'Gibt es Sozialprestige?' <u>Zeitschrift für Soziologie</u>, 14:

    209-235.

Zapf W. (1983): 'Entwicklungsdilemmas und Innovationspotentiale in modernen

    Gesellschaften', in Matthes J (ed): <u>Krise der Arbeitsgesellschaft</u>.

    Frankfurt: Campus Verlag: 293-308.

TABLE 1 *Approaches to the Time-Dependence of Career Opportunities and Hypothesized Effects of Variables*

Variable		Static approaches Status attainment theory	Semi-static approaches Human capital theory	Vacancy competition theory	Dynamic approaches
Time in labor force (life-course effect)	U D		–	–	(–)
Education	U D	+	+ –	+	(+) (–)
Prestige	U D		– +	–	(–) (+)
Level of modernization at entry into the labor market (cohort effect)	U D				– –
Labor market conditions at entry into the labor market (cohort effect)	U D				– –
Level of modernization (period effect)	U D				+ +
Labor market conditions (period effect)	U D				+ +

Key: U = Upward shifts, D = Downward shifts

TABLE 2 *Variables Used in the Analysis: Definitions, Means and Standard Deviations*[a]

Variable	Definition	German men Mean	S.D.
Time in labor force	Measured as the number of months from entry into labor force to start of job i.	228·00	130·90
Education	Highest grade completed at entry into the job i, measured in years:	11·21	2·05
	9 = Lower secondary school qualification (completion of compulsory education)		
	10 = Middle school qualification (certificate from *Realschule*)		
	11 = Lower secondary school qualification with additional vocational training (apprenticeship or certificate from specialized vocational school)		
	12 = Middle school qualification with additional vocational training degree (apprenticeship or certificate from specialized vocational school)		
	13 = *Abitur* (included in this category are certificates from a *Gymnasium*, *Kolleg* or *Wirtschaftsgymnasium*; also certificates from a secondary technical school, the *Fachoberschule* or the *Höhere Berufsfachschule*)		
	17 = Professional college qualification (certificate from a higher technical college or a professional college, the *Fachhochschule, Ingenieurschule* or *Höhere Fachschule*)		
	18 = University degree (from all institutions of higher education)		
Prestige	Wegener (1985) prestige score for job i.	54·93	22·53
Level of modernization at time of entry into the labor market	Factor score of the factor 'level of modernization' at entry into the first job.	–0·88	0·63
Labor market conditions at time of entry into the labor market	Factor score of the factor 'labor market conditions' at entry into the first job.	–0·94	1·62
Level of modernization	Factor score of the factor 'level of modernization' at time t.	0·29	0·96
Labor market conditions	Factor score of the factor 'labor market conditions' at time t.	0·04	1·02

Note: (a) Reported means and standard deviations are for sub-episodes, not individuals.

TABLE 3   *Results of the Factor Analysis*

Indicators of the development of social and economic structure	Initial factor matrix (using principal factoring without iteration)			Equimax rotated factor matrix	
	Factor 1	Factor 2	Communality	Factor 1 'Level of modernization'	Factor 2 'Labor market conditions'
Level of productivity[1] 1950–82	0·99103	0·13107	0·9993	0·9963	0·0824
National income *per capita* (deflated)[2] 1950–82	0·99135	−0·05155	0·9854	0·9876	−0·0999
National income per economically active person (deflated)[2] 1950–82	0·99016	−0·08248	0·9872	0·9849	−0·1308
Private consumption (deflated) 1950–82[3]	0·98891	−0·07543	0·9836	0·9840	−0·1237
Proportion of expenditure on services in private consumption (1950–82)[4]	0·93993	0·32645	0·9900	0·9548	0·2801
Proportion of gainfully employed in public sector (1950–82)[5]	0·99337	−0·06354	0·9908	0·9891	−0·1121
Proportion of 13-year-old pupils attending 'German Gymnasium' (1952–82)[6]	0·98084	−0·19544	1·0002	0·9701	−0·2432
Proportion of gainfully employed in service sector (1950–82)[7]	0·99406	0·07955	0·9944	0·9967	0·0308
Proportion of students in resident population (1950–82)[8]	0·98231	−0·11428	0·9780	0·9756	−0·1622
Proportion of civil servants in economically active population (1950–82)[9]	0·95612	0·10148	0·9244	0·9599	0·0546
Proportion of white-collar employees in economically active population (1950–82)[9]	0·98564	0·15581	0·9958	0·9921	0·1074
(Negative) Unemployment rate (1950–82)[10]	0·12728	0·95142	0·9214	0·1737	0·9441
Proportion of gross national product going to investment in plants and equipment (1950–82)[11]	−0·14626	0·94295	0·9105	−0·1000	0·9490
Proportion of registered vacancies of all dependent jobs (1950–82)[10]	−0·19634	0·89208	0·8343	−0·1525	0·9006
Eigenvalue (variance accounted for by factor)	10·6704	2·8252			
Proportion of total variance accounted for by factor	76·2	20·2			
Cumulative proportion of total variance accounted for	76·2	96·4			

Sources:

(1) 1950–60 Jahresgutachten des Sachverständigenrates zur Begutachtung der gesamtwirtschaftlichen Entwicklung 1967/68, p. 301
1960–82 Jahresgutachten des Sachverständigenrates zur Begutachtung der gesamtwirtschaftlichen Entwicklung 1983/84, p. 307
(2) 1950–59 BMWI, Leistung in Zahlen '67, p. 17
1959–74 BMWI, Leistung in Zahlen '75, p. 19
1974–82 BMWI, Leistung in Zahlen '82, p. 13
(3) see (1), 1967/68, p. 212 and.1983/84, p. 308
(4) 1950–60 Jahresgutachten des Sachverständigenrates zur Begutachtung der gesamtwirtschaftlichen Entwicklung, 1966, p. 208–209
1960–72 Jahresgutachten des Sachverständigenrates zur Begutachtung der gesamtwirtschaftlichen Entwicklung 1972/73,

p. 220–221
1972–82 Jahresgutachten des Sachverständigenrates zur Begutachtung der gesamtwirtschaftlichen Entwicklung 1983/84, p. 320
(5) Wirtschaft und Statistik 6/1985, p. 243
(6) Köhler (1977), p. 54 and Köhler (1986), p. 65
(7) 1950–60, see (4), 1972/73, p. 205
1960–82, see (1), 1983/84, p. 295
(8) 1950/51–1966/67, BMWI, Leistung in Zahlen '67, p. 29
1966/67–1970/71, BMWI, Leistung in Zahlen '72, p. 31
1970/71–1982/83, BMWI, Leistung in Zahlen '82, p. 26
(9) 1950–60, Statistisches Jahrbuch 1950, Statistisches Bundesmat, Fachserie A, Reihe 6
1960–82 cf. Statistische Jahrbücher 1961 to 1983
(10) see (1), 1983/84, p. 294
(11) 1950–60 see (4), 1972/73, p. 214
1960–82 see (1), 1983/84, p. 310

TABLE 4 *Estimates of Models for Transition Rates to Better Jobs, Similar Jobs and Worse Jobs for German Men Born 1929–31, 1939–41 and 1949–51*

Estimates for model	Upward moves 1	2	3	4	5	Lateral moves 1	2	3	4	5	Downward moves 1	2	3	4	5
Log of mean rate	−6·135					−5·321					−6·352				
Constant		−4·241***	−4·288***	−4·943***	−5·585***		−3·388***	−3·148***	−3·669***	−4·047***		−4·413***	−1·893***	−2·928***	−3·377***
Time in labor force (life-course effect)		−0·012***	−0·015***	−0·012***	−0·082***		−0·012***	−0·012***	−0·013***	−0·076***		−0·012***	−0·013***	−0·014***	−0·086***
Education			0·187***	0·224***	0·268***			−0·023	0·005	0·036			−0·354***	−0·291***	−0·254***
Prestige			−0·042***	−0·041***	−0·042***			0·001	0·001	0·001			0·026***	0·027***	0·025***
Level of modernization at entry into labor market (cohort effect)				−0·294**	−9·664***				−0·326***	−8·756***				−0·520***	−10·230***
Labor market conditions at entry into labor market (cohort effect)				0·009	−1·394***				0·052	−1·259***				0·039	−1·311***
Level of modernization (period effect)					9·066***					8·300***					9·335***
Labor market conditions (period effect)					1·203***					1·133***					1·150***
Number of shifts	590	590	590	590	590	1332	1332	1332	1332	1332	475	475	475	475	475
Number of sub-episodes	22843	22843	22843	22843	22843	22843	22843	22843	22843	22843	22843	22843	22843	22843	22843
x^2		842·21***	1007·46***	1024·02***	2165·29***		1980·16***	1982·53***	2016·12***	4312·67***		711·11***	827·84***	863·44***	1763·86***
df		1	3	5	7		1	3	5	7		1	3	5	7

* Statistically significant at 0·05 level; **statistically significant at 0·01 level, ***statistically significant at 0·001 level.
Rates are measured with month as units.
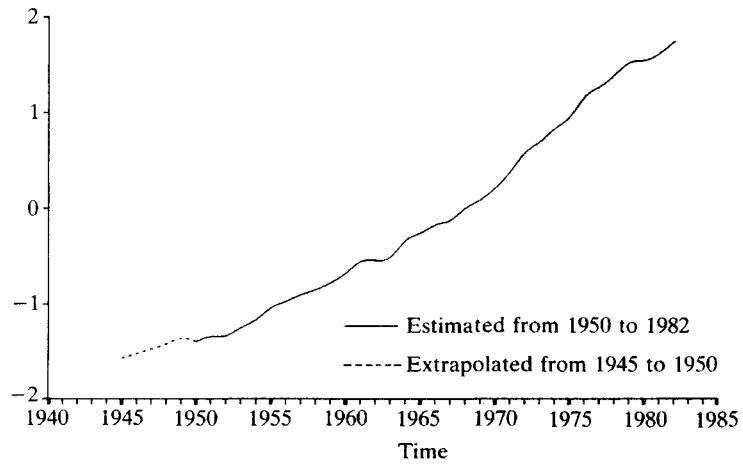
FIGURE 1   *Development of Modernization*
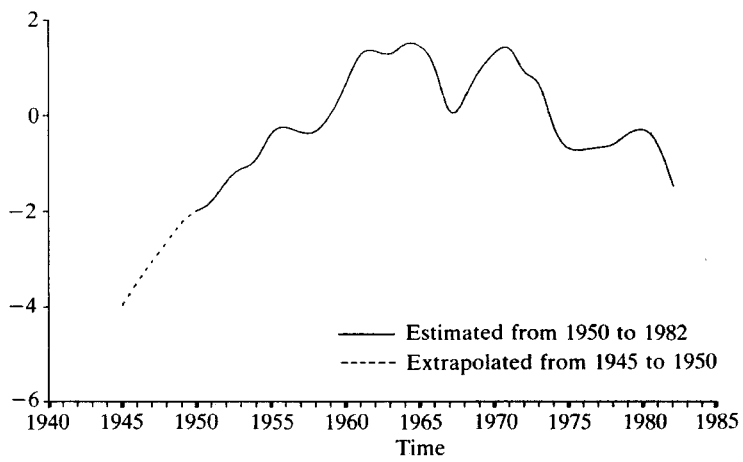
FIGURE 2   *Development of Labor Market Conditions*
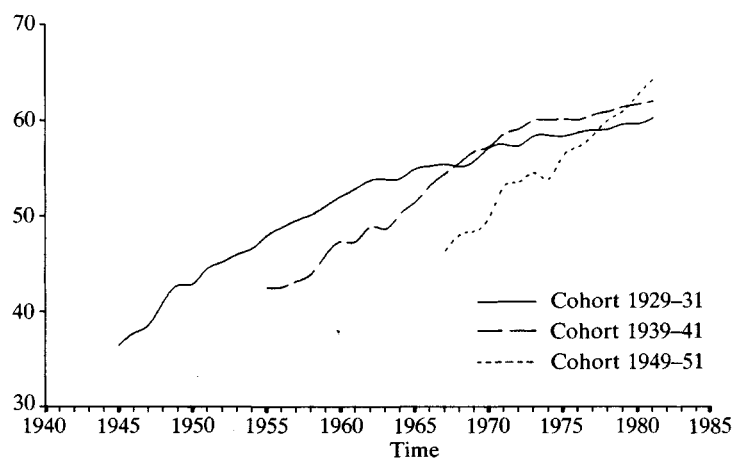
FIGURE 3   Development of Prestige Scores for Men over Historical Time
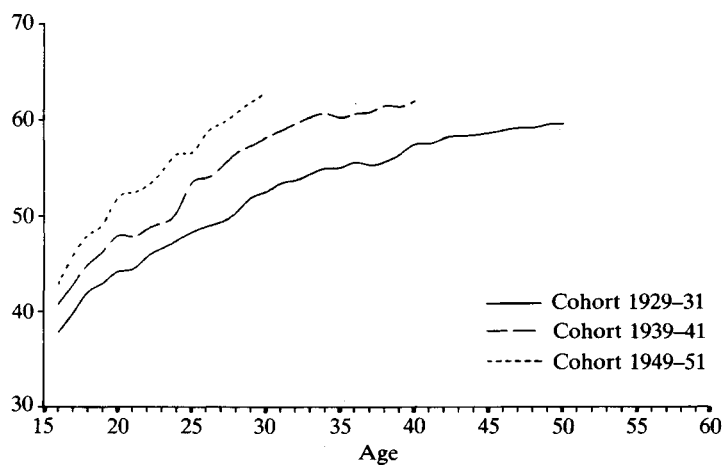
FIGURE 4   Development of Prestige Scores for Men over the Life-Course
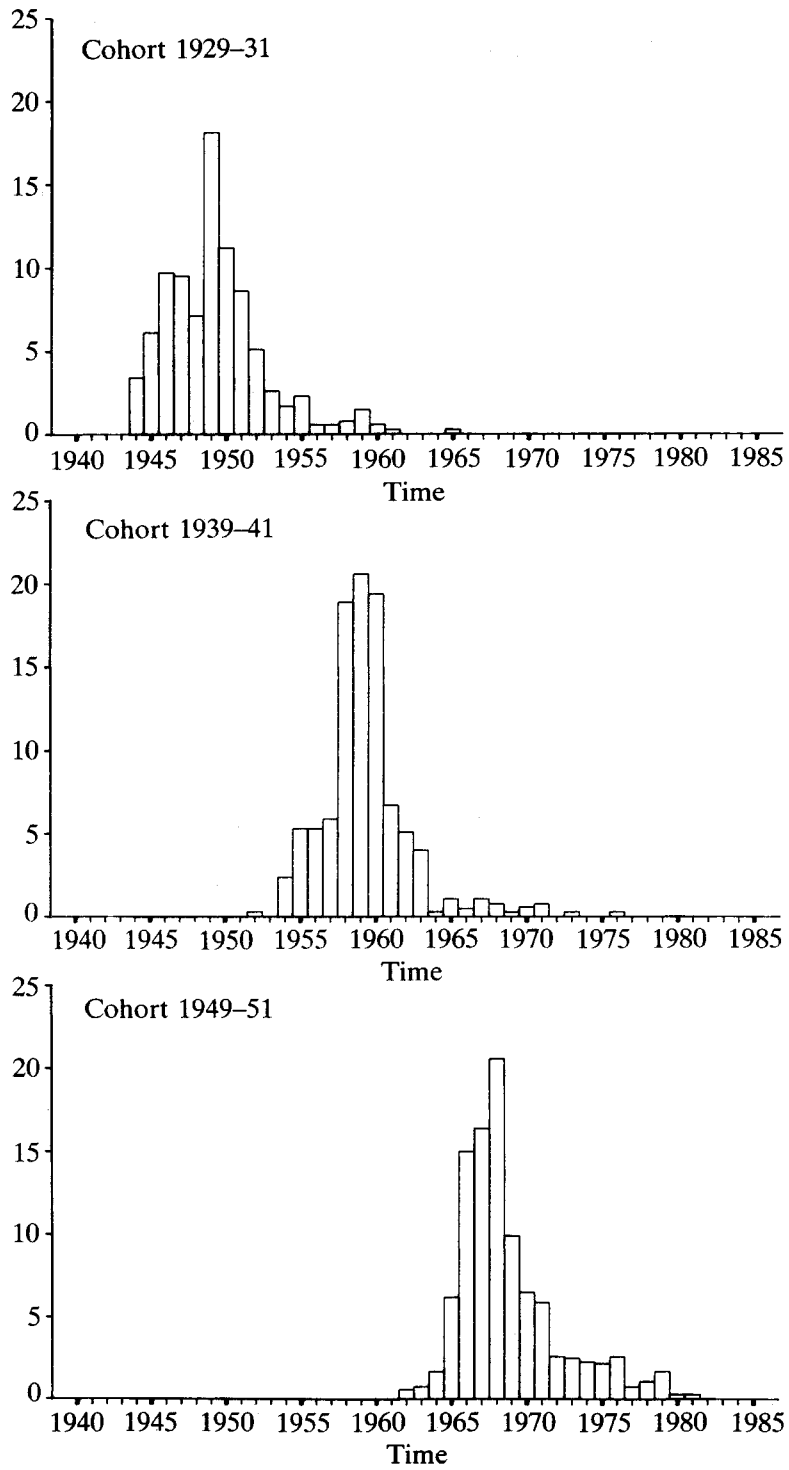
FIGURE 5   *Distribution of Entry into the Labor Market*

Job-shifts in the career beginnings of Israeli men

Yossi Shavit, Judah Matras, and David L. Featherman

INTRODUCTION

The topic of this paper is at the intersection of two
research domains: (1) the study of career beginnings of
young men--whether along or as part of "transitions to
adulthood," and (2) the study of labor force participation,
employment, and unemployment of young men.

Events in the early stages of careers are important
determinants of socioeconomic attainments in subsequent
phases of the life course. Models of socioeconomic
attainment (e.g. Blau and Duncan, 1967; Featherman and
Hauser, 1978) demonstrate considerable stability in
occupational prestige between jobs held early and late in
the life course. Indeed, much of the importance of education
in the attainment process is mediated by its effects on

occupational allocation early in life. Hogan (1982) demonstrates that the temporal ordering of entry into the labor force, school leaving, and other such milestones in the transition to adulthood also affects ultimate attainment. Coleman (1984) argues that some of the occupational advantage of Whites over Blacks in the United States is due to the fact that Whites are more likely to accumulate work experience while still in school and to return to school after initial phases of employment (see also Freeman and Wise, 1982). Osterman (1980) shows that job instability in late adolescence is detrimental to orderly career formation in early adulthood. In short, the process of entry into the labor force seems to be an important nexus in the socioeconomic life course and worthy of the attention it has been receiving (Ornstein, 1976; Osterman, 1980; Spenner et al. 1982).

But highly industrialized and developing countries alike have in this century experienced a trend toward diminished labor force participation among young adolescent and early adult males. Indeed, in most countries a substantial majority of persons under 20 years of age are not in the labor force at any moment in time. Some have not yet entered the labor force for the first time while others have withdrawn temporarily (Hauser, 1979). For those adolescents who are out of school, labor force participation is erratic and unrewarding. (e.g. Osterman, 1980; Mare et al., 1984; Coleman, 1984; Borus, 1984; Matras et al., 1984). When compared to adults, youths are more likely to be unemployed

(e.g. Mare et al., 1984), to hold jobs for shorter periods of time (e.g. Osterman, 1980), and to work in low paying, dead-end type jobs which are abundant in the secondary labor markets (e.g. Brown, 1982). In short, youths are said to be a marginal group in the labor force and to suffer from handicaps similar to those suffered by ethnic and racial minorities and by women (Osterman, 1980).

One cause for the marginal status of adolescents in the labor market is their higher rate of job turnover. Job changes often involve spells of frictional unemployment. Furthermore, as persons leave jobs, they often fail to realize investments of time and on-the-job training.

In this paper we develop a model which accounts for age variations in job instability among Israeli workers. We have three objectives:

(1) to continue, add to, and elaborate our study of a cohort's transition to adulthood in Israel. We have been studying and measuring various aspects of the 1954 Israeli birth cohort's (a) post- primary and post-secondary schooling, informal education, and school- leaving; (b) military service; (c) employment, including occupational attainment, income, job authority, and job satisfaction; (d) household formation, including departure from parental home, marriage and parenting, and acquisition of own residence. Ours is the first such study in Israel.

(2) to contribute to the description and discussion of factors in employment and non-employment of young Israeli men, especially at the early post-military-service stage.

This is an area in which very little has been done or undertaken in Israel though increasingly perceived as one demanding attention. Although there are census and survey data available on labour force participation, employment, and unemployment, virtually none of the kinds of studies carried out in the United States, Germany, or the United Kingdom of youth employment or non-employment has been replicated in Israel.

(3) to contribute to the general discussions of youth employment and career beginnings in the transition to adulthood; and, in particular, to study durations of jobs, job spells, job tenure, or job stability and instability--a topic which intersects the two domains: career beginnings, and employment and non-employment, in adolescence and early adulthood. Despite the vast and impressive literature, not very much of it addresses the question of employment stability using full employment histories and in conjunction with full histories in other life domains. We hope to begin to do so with our data.

INSTABILITY OF YOUTH EMPLOYMENT

Spilerman's (1977) work on careers provides useful conceptual tools with which to analyze the dynamic aspects of labor force participation (Spenner et al., 1983). Careers are defined as structured sequences of work roles which persons occupy over time. Careers vary in the level of occupational rewards which their constituent jobs bestow upon workers at any point in time. They also vary in the

extent of continuity which they involve. Some persons embark upon careers which involve numerous job changes while others remain in a single job for a long time. Some careers link progressively more rewarding jobs while others manifest an erratic pattern of change in reward level. Spilerman distinguishes between _orderly_ and _chaotic_ careers. The former refers to job sequences which involve a consistent improvement in reward level while the latter refers to sequences in which a unilineal process is absent.

Frequent shifts among jobs can be either beneficial or detrimental to socioeconomic attainment (Spilerman, 1977; Berg, 1970; Coleman, 1984). Their benefit lies in the fact that movements are necessary if upward mobility is to take place. However, rapid shifts may also reflect inability to accumulate on-the-job training and experience, or to enjoy promotion opportunities within,the firm. But being employed at any moment in time, or over a time span, is a consequence of a sequence including (all of): i. seeking employment, or, at least, readiness to accept a job; ii. finding a job, or having a job offered; and iii. keeping or staying on the job.

Not being employed at any moment in time, or being employed not at all or very little over a time span, is a consequence of ANY of: i. not seeking a job, or non-readingess to accept a job; OR ii. seeking but not finding, or not being offered, a job; OR iii.,finding or having a job but not keeping, or not staying on the job.

So job stability is positively associated with the probability of employment, or with duration of employment, or with proportion or fraction of total life years employed. Job instability is always associated with the probability of non-employment; but there is also non-employment independent of job durations or stability.

Increasing job stability over time or over successive ages in the life course implies an "orderly" career given, also, consistent improvement in reward level. Job instability over time or successive ages implies maladaptation in the labor force and is detrimental to socioeconomic achievement. But early worklife job instability associated with upward moves may imply "ordered" careers, provided it is succeeded ultimately by increasing job stability.

There is considerable indirect, and some direct, evidence of a positive relationship between age and job stability. Adolescence is said to be a period of emotional and social moratorium during which people try out various social roles and refrain from long term social obligations (Erikson, 1968). Most adolescents are still not married nor are they yet expected to support their parents. Consequently, they do not value a steady income as much as older workers do. In the labor force this is reflected in high turnaround of youths among jobs and in repeated entries into and exits from the labor force. The transition into adulthood involves a gradual change into a more stable pattern of employment. Frequent shifts may be beneficial in adolescence when young workers are still searching for ever better jobs. An orderly

career, namely one which leads the incumbent to social success, exhibits a gradual decline in the rate of job changes: As workers grow older and accumulate investment in the work place, they become less likely to risk fresh starts. With age, instability of employment comes to signify a maladaptation to the demands of the labor market and is progressively detrimental to socioeconomic attainment.

The differential preferences of young and older workers has been asserted to be compatible with the differential needs of firms in the economy (Osterman, 1980). Some firms require unstable employees. These are firms which are typically characterized as belonging to the secondary labor market. They are often small firms in which the production process is labor intensive and in which little training of workers is required. Adolescents are said to be suitable workers for such firms: On the one hand they do not value promotion opportunities or job security and on the other hand they can easily get access to jobs in the secondary economy where prior experience is not required. By contrast, primary labor market firms are typically large, capital (or human capital) intensive, and require workers who are willing to undergo training. Since training of workers is costly, primary sector firms attempt to employ older workers who are likely to stay with the firm for long periods of time (Thurow, 1975). Similarly, within the various industrial sectors in the economy, some occupations and jobs require stable workers whereas other jobs are more 'suitable' to casual labor.

Osterman argues that the compatibility between the needs of firms and jobs, on the one hand, and the social-psychological nature of age groups, accounts for the differential concentration of age groups in occupations and industries. However, the sectoral concentration may have a reciprocal causal relationship with youth employment patterns. Brown (1982) for example, suggests that youth unemployment is due in part to the fact that adolescents are over-represented in occupations which are more strongly associated with spells of unemployment. Similarly, the instability of adolescent workers may be due to their concentration in industries which do not provide continuity of employment within firms and jobs. Clearly, the two explanations are not mutually exclusive: adolescents' patterns of employment may result from both factors (i.e. from the nature of adolescence and from the nature of the jobs they hold). One objective of the present analysis is to contrast the magnitudes of the two effects in explaining age variations in job instability.

Two additional factors single out youth from adult workers. First, youths have not yet had the opportunity to make substantial investments in Human Capital. Wages, earnings, job stability, and employment are positively related to educational attainment and to experience in the labor force (e.g. Mincer, 1974; Featherman and Hauser, 1978; Hachen, 1983). Educated workers are perceived by employers to be more productive. Consequently, such workers manage to secure entry into the more rewarding jobs which, in turn,

they are less likely to leave (Carroll and Mayer, 1985).
Furthermore, employers are said to invest more heavily in
the training of educated workers for the requirements of the
job. Once the investment has been made, trained workers are
less likely to be fired (Thurow, 1975). Many adolescents
have not yet had the opportunity to complete their course of
studies and many enter the labor force while still in school
or in between spells of schooling (e.g. Coleman, 1984).
Thus, adolescents may simply exhibit labor force behavior
which is characteristic of less educated workers in general.
Similarly, adolescent workers have not had the opportunity
to accumulate experience in the labor force. In the eyes of
employers, experience, like formal schooling, is an
indicator for potential productivity on the job (e.g.
Mincer, 1974; Bills, in progress). Furthermore, experienced
workers are more likely to have completed their search for a
career, to 'settle down' in jobs, and to leave them less
readily (Carroll and Mayer, 1985).

Second, working youths are a negatively selected group
from amongst each cohort. As noted earlier, in
industrialized societies schooling is normatively extended
later into adolescence and early adulthood. Mare and his
associates (1984) suggest that as education is historically
prolonged, only the least able of each successive cohort
ever experience fulltime work before age 18. They argue that
youth unemployment is due in part, to the fact that the pool
of working youths consists primarily of persons who are
least able (for otherwise they would have still been in

school). The adult labor force on the other hand, consists of a more representative subset of the ability distribution. Thus, with age, as more able persons leave school and enter the labor force, the modal pattern of careers becomes more orderly, stable, and rewarding. This intriguing hypothesis is consistent with findings reported by Mare and his associates on rates of employment in the transition from youth to adulthood. However, the data which they employ do not include measures of ability. Consequently, they test the hypothesis indirectly within the framework of a statistical model of selection bias. Our data enable us to test for selectivity effects on job instability in adolescence.

In summary, the labor market experience of youths is said to display higher rates of job turnover. Four factors are said to account for these phenomena: the socio-psychological nature of adolescence, the industrial and occupational concentration of youthful workers, the human capital profiles of young workers, and their negative selection from among the total cognitive ability distribution. The relative importance of these effects in Israel are the concern of the present study.

EMPLOYMENT AND THE TRANSITION TO ADULTHOOD IN ISRAEL

Most Israelis spend about three of their late adolescent years in military service. The near-universal and lengthy service has several implications for the process of career

formation. First, a large part of the population is kept out of the labor force, thereby restricting an over-supply of labor, especially of those age groups of workers who in other societies suffer relatively severe unemployment. Thus in Israel military service may effectively 'substitute' for youth unemployment.

Second, for an unknown but probably significant proportion of each cohort, military service constitutes an opportunity to acquire some form of on-the-job training. Many soldiers participate in courses which provide skills that are transferable to civilian jobs (e.g. drivers, cooks, electronic technicians). Third, because of its near-universality, military service has acquired the status of a major rite de passage. It is perceived as a major demarcator between youth and adulthood. Thus, it is hypothesized that post-military age workers are more stable. On the other hand, in the pre-military ages, working youths realize that their spells of employment are soon to be interrupted by military conscription at age eighteen. If jobs are viewed as temporary, they may be discarded more readily.

In Israel there has been remarkably little systematic study of labor force and employment patterns of adolescents. Studies of employment and its alternatives among adolescents and young adult men have addressed primarily the issues of causes and educational outcomes of dropout, especially from secondary education (e.g. Shavit, 1984) and occupational outcomes of vocational education (e.g. Starr and Kahane,

1984; Nahon, 1984). The publication of extensive or detailed census statistics on age and employment is limited because the Israeli Central Bureau of Statistics does not publish data for detailed age categories. Thus, features of the Youth Labor Market in Israel have not been described or analyzed in depth. Indeed, its very existence, as distinct from the adult labor market, has not been demonstrated.

Recently the rate of youth unemployment in Israel has grown to Western-like proportions. For the first time in nearly twenty years unemployment is considered a major social problem. In the mid-1980s the overall unemployment rate approached nine percent and the unemployment among adolescents and young adults approached 20 percent. Thus, it is high time for a description and analysis of the processes by which young Israelis make the transition from youth to adult patterns of employment.

We consider first the age pattern of employment among males of the 1954 birth cohort as its members moved from age fourteen to age 26 (The data will be described in detail below.). The percent employed (whether full or part time) at the beginning of each half-year age interval: 14, 14.5, 15... for ages 14 to 26 is plotted in Figure 1. The patterns are plotted separately for persons of Asian-African origins (hereafter AA) and for those of European-American birth or ancestry (EA)(1).

---

1. The distinction between Asian-Africans and European-Amercians is a major basis of social differentiation in Israel (e.g. Smooha, 1978). It is highly

----
Figure 1 about here

About 14% of the cohort reported employment at age fourteen. This percentage increased sharply among the AA youths to a peak of 49 percent at age 17.5, and increased more gently for the European- Americans to a peak of 35 percent at that age. There is a sharp drop - reaching a trough of 12 percent - in the percent employed at ages 18, 19, and 20, when most of the young men were performing compulsory military service. After the military service ages, the percent employed recovered very dramatically, reaching about 85 percent by age 22. Subsequently there was a less dramatic but steady increase in the percent employed reaching 90 percent at the ages 24 and 25. The major difference between the ethnic groups is in the much larger proportion of AA's who worked prior to military service. This probably reflects their lower rates of secondary-school attendence (e.g. Shavit-Streifler, 1983). Yet, and despite documented ethnic differences in the attendence of higher (post-military) education, there was but a negligible difference between the groups in employment in the ages 20 through 26. This may reflect the fact that Israeli students in post- secondary education typically work, at least in part-time jobs.

correlated with all indicators of socioeconomic attainment. We present the patterns separately for the two groups because it would be unacceptable to assume ethnic homogeneity of the patterns.

Figure 2 about here

In an earlier report on the 1954 cohort note was taken of
the intensity of employment commitments in early
adolescence. Specifically, it was reported that a large
proportion of jobs taken prior to military service were in
fact full time jobs and held for at least five months
(Matras et al., 1984). In Figure 2 we plot the mean weekly
hours of employment in jobs held by the cohort at each of
the ages fourteen to 26. Among EA's there was a rise, at the
ages of military service, from a level of about 37 hours per
week, to a level of nearly 50 hours. Among the AA's on the
other hand, the change is less abrupt: Through-out the age
range, Asian-Africans worked, on the average, full-time or
more. Thus, not only were the percentages employed in
adolescence sharply contrasted among the groups, but also,
the work commitment of those employed differed: Age
variations are more pronounced in the case of the EA's. It
would seem that the prolonged schooling of this group
accounts for their lower employment rates in adolescence and
for their somewhat more partial mode of employment in those
ages. The most interesting finding in Figure 2 is that,
even in adolescence, in both groups those who worked did so
a large number of hours per week.

DATA

The data which we employ consist of 2144 retrospective
life histories obtained in interviews with a national
probability sample of Jewish Israeli men who were born in
1954 (Matras et. al, 1984). The interviews were conducted
during the years 1980-81 when the respondents were about 27
years old. The interviews consisted of two sections: (i)
Retrospective life histories on various areas of activity,
and (ii) Details of current employment, family
characteristics, social participation, attitudes, and
income. The life history section of the interview schedule
reconstructs residential, educational, employment, family
formation, and military service histories. Within each of
these domains of activity, the respondent was asked to list
each event, to describe it in some detail and to date it.
For example, in the domain of employment, the following
information is available on each job held since age
fourteen: Date of entry and exit, occupational and
industrial codes, number of hours worked per week, number of
subordinates, whether or not Respondent changed a job-title
or section within the firm, and dates of changes. In
addition, information is also available on whether or not
the job change involved a change of employer. Retrospective
data on income were not collected because at Israel's high
inflation rates, respondents were often unable to recall
either nominal or real income figures.

The interview data were merged with military, school, and
police records which were retrieved from the respective

government agencies. These external data sources provide information on measured intelligence at ages 13 and 17, arrest records, school grades, and various characteristics of military service histories. The present analysis employs subsets of the interview and military data(2).

We define a job as a spell during which a person is employed by a given employer. As of the time of interview, the 2144 members of the sample had had some 8222 distinct and identifiable jobs. In about seven percent of these jobs, the respondent was self employed. These jobs are deleted from the analysis because the processes which determine the likelihood to 'shift' from self-employment are probably very different from those which determine shifting from jobs in which the worker is an employee. One hundred and thirty jobs were entered before age fourteen. These are deleted because in the interview schedule, respondents were asked to list jobs which were undertaken after that age. Also deleted are all jobs which were held prior to Respondent's immigration to Israel (155 jobs). Of the remaining 7500 jobs, two-thirds where entered at the post-military-service ages (21-27). About one-fourth (22.7 percent) were entered by age eighteen - that is, at the post-primary school ages. Finally, 10.7 percent were jobs entered at the military service ages (18-20).

2. See the following for analyses which employ other subsets of the data base: Shavit, 1984; Shavit and Featherman, 1985; Shavit and Rattner, 1986.

The multivariate analysis which is reported below employs Partial Likelihood Proportional Hazard models which are estimated in SAS's PHGLM procedure (Harrell, 1982). With 7500 spells, the estimation proved very time consuming. In order to save computer time, a 33 percent, random, representative sample of respondents was drawn from the original sample of 2144 persons. The sub-sample consists of 738 respondents who were employed in 2448 jobs which constitute 32.4 percent of the jobs which had been previously selected.

THE MODEL

At the outset we reviewed four explanations for age group differences job stability. The four accounts are contrasted within Proportional Hazards (Cox, 1972) models of the rate of job shifts. Proportional Hazards models are described extensively by others (e.g. Tuma and Hannan, 1984; Allison, 1984). Briefly, the model specifies the natural logarithm of the rate to be

$$\ln(r_t) = b_1 X_1 + \cdots + b_k X_k$$

where r_t is the hazard rate with which person i is expected to leave a job at some small time interval. The X's are independent variables (e.g. age group, occupational prestige, marital status) which affect the hazard, and the b's are parameters to be estimated. Positive b's indicate

that the corresponding independent variable increases the rate, i.e., that it increases a worker's 'tendency' to leave a job at any point in time. A negative value of the coefficient indicates that the corresponding X reduces the tendency to leave a job, i.e., that it increases job stability.

The units of analysis are job spells, i.e., records which pertain to jobs held by individuals. Each spell is characterized by values on nine sets of variables. The first is the duration of employment of the respondent on the job measured in months. This variable provides the basis for calculating the rate of leaving the job. Longer durations reflect smaller hazards because, by definition, persons with high hazards of leaving jobs are unlikely to remain in them for long durations. Duration increases monotonically with age from 12.03 to 13.85 (see Table 1). Second is a variable indicating whether the job was censored by the time of interview. Of jobs entered in adolescence, only one percent were still held at the time of the survey, compared to 35 percent of jobs in the older age groups. Third, are dummy variables representing the age category of job entry.

Fourth, is a set of variables which characterize the nature of the job to which the spell pertains. Six job characteristics are included in the analysis. PRESTIGE is the prestige score of the job title (Kraus, 1983). This variable ranges from 3 to 98. It is entered into the model as a proxy for the overall 'desirability' of the occupation. We hypothesize that workers are less likely to leave 'good'

jobs. Thus, we expect the effect of PRESTIGE on the hazard to be negative. PUBLIC is a dummy variable coded 1 if the job was held in one of several industries which are monopolized (or nearly monopolized) by the State, local authorities, or other public agencies (e.g. education, health, welfare, religious services, etc..). About 30 percent of the Israeli labor force is employed in the public sector. One important characteristic of public employment is that it assures many workers job security and provides promotion opportunities within the various agencies and public firms. Thus, we expect that the hazard rate of leaving public sector jobs is lower than the hazard rate of leaving other jobs. NSUBS is the number of workers who where subordinate to our respondent on that job. We expect that workers who are in positions of authority had accumulated on-the-job training in the firm and are more likely to be retained by their employers. HOURS is the number of weekly working hours. Full time workers are less likely to leave jobs than part time workers. Thus, the effects of NSUBS and HOURS on the hazard are expected to be negative. These six job characteristics are expected to account for some of the age differences in job stability because younger workers are less likely to be employed in the public sector, but are more likely to hold low prestige occupations, to have little job authority, and to work part time (Table 1).

Table 1 about here

A fifth set of variables measures human capital at the time the worker entered the job. Older workers have had time to accumulate more work experience and education than young workers. We hypothesized that job stability is positively affected by the educational credentials which workers bring to the job. Educational credentials are entered into the model as four dummy variables which correspond to some or complete SECONDARY education; secondary education plus a MATRICULATION diploma; some UNIVERSITY education; and a university DEGREE (a B.A. or higher). The omitted category is 'no secondary education'. EXPERIENCE measures the number of months of labor force experience that the worker brought to the current job. Older workers have had more lifetime in which to accumulate labor force experience. We expect that with time in the labor force workers become more committed to their jobs and exhibit more job stability.

The sixth set of variables consists solely of Respondent's intelligence (IQ). The variable is measured by the military screening examinations which most Israelis are required to take at age seventeen. The test consists of a Raven (Raven, 1958) matrices test of analytic intelligence and of a test of verbal intelligence. (See Shavit and Featherman, 1985 for a description of the military aptitude tests.) The military test scores were merged into our data for all of the respondents. As we noted earlier, it has been argued (Mare et al., 1984) that adolescent workers are typically less intelligent than adult workers. From the appropriate row of Table 1 we learn that the mean intelligence score of

those holding jobs entered in adolescense (14-17 years) is
51.65 whereas the mean IQ scores of those with jobs entered
at the two later age stages are 54.97 and 57.63
respectively.

The seventh set of variables includes three dummy
variables - STUDY, MARRIED, and CHILD - which indicate
whether or not Respondent was in school, married, or a
parent at the end of the job spell in question. We expect
the effect of school enrollment on the hazard to be positive
because students are more likely to work in temporary
(summer) jobs. The effects of marital status and being a
parent are expected to be negative because workers who must
support families are less likely to shift among jobs.
Controlling for these variables is expected to attenuate age
group differences in job stability.

The eighth set of variables which characterise each job
spell are two indicators of worker's socioeconomic and
ethnic origins. POPSEI is the prestige of father's
occupation when Respondent was a teenager, and ETHNIC is a
dummy representing Asian-African ethnic origins. Adolescent
workers are disproportionately of AA origins (because the
school drop-out rates of this group is larger,
Shavit-Streifler, 1983) and of low socioeconomic background.
To the extent that these two variables affect job stability,
they must be controlled in our models.

A final set of control variables which we include in our
models consists of dummy variables representing the
sequential number of the job in the Respondent's work

history. Nine such dummies are included representing the first job, the second job, etc.. through the ninth job. The omitted category pertains to the 'tenth or more' job. The sequential position of the job in the work history is controlled because 'unstable' workers change jobs more rapidly and contribute more observations to our sample of spells. Thus, there is an over- representation of short, unstable jobs in our sample. To the extent that the over-representation of unstable jobs is more prevalent in one of the age groups, the age effect on job instability may be biased. Indeed, the mean sequential position of jobs is positively related to age of entry (See the means of the sequential number of spell, in Table 1). By controlling for sequential position of the job, we eliminate this bias.

AGE DIFFERENCES IN JOB STABILITY

In Figure 3 we present the adjusted cumulative proportions of jobs by durations. The adjustment is carried out in the context of Survival Analysis. The analysis is conducted for jobs entered at each of the three age groups (ages 14-17, 18-20, and 21-27). A job is defined as censored if it was still held by Respondent at the time of interview.

Figure 3 about here

The horizontal axis of the figure corresponds to the number of months a person 'survived' in the job. The vertical axis corresponds to the proportion of the group who survived to a given duration. The median duration for the

jobs begun at adolescent, military and adult age groups are 11.63, 12.46, and 13.79 months respectively. The somewhat flatter slope of the "adult" curve (i.e., jobs begun at adult ages) suggests that adults are somewhat more stable workers who are more likely to persist in jobs. The job stability at the other two age stages appears very similar.

Table 2 about here

In Table 2 we present estimates of Proportional Hazards Models of job leaving. Model 1 indicates that in the older of the three age groups, Respondents were least likely to leave their job at any point in time. The effect of the 18-20 age stage is small and not significantly different from the effect of the omitted age category (14-17). This pattern is consistent with the finding reported in Figure 3. Model 2 estimates age effects which are adjusted for the sequential position of the job in the work history. When compared with Model 1 the adjusted effects of age indicate a slight accentuation of the differences between jobs held at the oldest and youngest ages. Model 3 is our baseline model in that it adjusts the age effects for both sequential position of the job and for Respondent's socioeconomic background. Under this model jobs which were entered by Respondents in the two older age groups were more stable than adolescent jobs.

Models 4 through 7 correspond to the four explanations for age differences in job stability which were discussed earlier. Model 4 tests the hypotheswas that adolescent

instability is due to the kind of jobs in which adolescents are employed. As expected, the rate of leaving a job was inversely related to occupational prestige and to job authority. Similarly, Respondents in the public sector persisted longer in their jobs than those in the private sector. Contrary to our expectation however, the inclusion of the four job characteristics in the model fails to attenuate the age effects substantially.

Model 5 indicates that higher education and labor force experience are conducive to stable employment but that the inclusion of these variables in the model also fails to explain the age effects. In Model 6 we control for Respondent's intelligence. Contrary to our expectation, the effect of this variable is positive and its inclusion in the model accentuates the age effects rather than attenuating them.

In Model 7 we control for the three life-course variables. As expected, Respondents employed while students were relatively unstable workers; while those employed as married men and parents were more stable than single workers. Furthermore, the net effects of age in Models 7 are about half their size in Model 3.

We conclude that of the four sets of explanatory variables, the major source of age-groups differences in job stability are age-related differences in school enrollment, marital status, and parenting. This conclusion is also supported by a comparison of the age effects in Models 8, 9, and 10: Model 8 controls for the four job characteristics,

the human capital variables, and for cognitive ability. The
net effects of age in this model are only slightly smaller
than those in the baseline model (Model 3). When student
status is added to the model (Model 9), the effects of age
are reduced by 30 and 50 percent. When marital status and
CHILD are added (Model 10) the age effects are rendered
insignificant.

The net effects of most control variables in Model 10 are
very similar to their effects in the earlier models. The
only substantial differences are in the effects of the
education dummies: In the full model, their effects are
more pronounced than in Model 5, and assume a monotonic
pattern, with the size of the effect rising with educational
level.

In summary, the relative instability of the Israeli
adolescent employees seems due primarily to their statuses
in the life course domains of schooling, and family. Their
lower level of human capital and the characteristics of the
jobs they hold contribute only marginally to adolescent job
instability.

JOB SHIFTS; THE DESTINATIONS OF MOVES, AND UNEMPLOYMENT

We noted at the outset that a job shift can be both an
instrument for social mobility (when workers shift to better
jobs) or an indication for disorderliness of a career
segment. We also suggested at the outset that as workers
grow out of adolescence their labor market behavior becomes
more career oriented in the sense that they seek to improve

their positions by upward shifts. In terms of our model, this translates into the expectation that the rate of moving into a better job is greater for young adults than for adolescents and that the rates associated with unemployment or less desirable destinations are greater in adolescence.

Following Sorensen and Tuma (1981) and Carroll and Mayer (1985) we distinguish between upward, downward, and lateral job shifts. Upward shifts are defined as moves into jobs in which the occupational prestige is higher by five points or more than that of the job of origin. Downward shifts involve a loss of occupational prestige (of five points or more). Shifts into jobs with similar prestige scores are defined as lateral moves. In addition we also take cognizance of moves into non-employment. Among jobs begun in adolescence, only 24.8 percent were followed directly by other jobs (Table 3). In the first column of Table 3, the most prevalent destination is military service: Thirty four percent of job spells were terminated within two months of induction into military service, 25.1 percent were followed by spells of school enrollment, and 15.3 percent of spells were followed by three months or more of being 'unemployed'(3).

Table 3 about here
----

3. Conventionally, unemployment is defined as a state of seeking employment while being out of a job. Our data base does not contain information on job search behavior. Consequently we employ the term unemployment in reference to periods during which Respondent was not employed, was out of school, and not in military service.

Of jobs entered in the 18-20 age category, 27 percent were followed by spells of unemployment, 18 percent were followed by induction into military service, and 41 percent were followed by direct entry into other jobs. In the oldest category, 35 percent of jobs were censored by the interview, 14 percent were followed by unemployment, and 44 percent were followed by other jobs.

Table 4 about here

In Table 4 we present models of job leaving which are conditional on the destination of the move. The first model is similar to the last model of Table 2 but here a job is defined as censored not only by time of interview but also by military conscription. That is, all job spells which ended within two months of induction are defined as censored. Conceptually, the model corresponds to the notion that time of induction is determined by processes which are external to labor market processes. Some (if not many) adolescent jobs would have persisted longer if the worker was not obliged to enlist at a specific age. By defining conscription as a censoring event, we focus on those modes of job-terminations which are initiated by either the employee or by the employer (but not by military service). The estimates of the model (Model 1 of Table 4) indicate that, when we 'discount' job leavings which were due to conscription, Respondents appear as more stable workers (ceteris paribus) as adolescents than in the other two age groups. The effect of educational attainment is now more

pronounced than revealed by our earlier analysis, but the effects of the other variables are similar to their effects in Model 10 of Table 2.

In Model 2 of Table 4 we focus only on departures which led directly to other jobs. All departures which were not followed (within two months) by another job, are defined as censored. Among job shifters, the greater stability at adolescence is even more pronounced than in column 1 of the table (compare 1.038 and 0.728 to 0.477 and 0.381). It would seem that much of adolescent instability was associated with moves into destinations other than immediate, subsequent jobs (i.e. back to school). Interestingly, Respondents as older workers were more likely than as adolescents to move in all three directions (Columns 3,4, and 5), including into less prestigeous jobs. If careers become more orderly with age, we would have expected the propensity of workers to make downward moves to diminish with age. The results contradict this expectation. Rather, the findings suggest that younger workers were least likely to move, irrespective of destination.

Even among jobs which led into spells of unemployment (see footnote 3 for our operationalization of the term), the effects of the older age categories are positive and significant, though smaller than the comparable effects in Column 1. Thus, Respondents as young adults were more likely to shift into unemployment but less so than they are likely to make other types of moves.

SUMMARY AND DISCUSSION

Adolescent workers are said to occupy marginal positions in the labor force: Their income and occupational prestige are low, they are more frequently unemployed, and they are less likely to benefit from promotion opportunities within and across firms. The marginal position of youths has been attributed in part to their erratic pattern of labor force participation and job instability.

In the present study we analyzed work-history data for a sample of young Israeli men in an attempt to describe and explain differences in job stability between adolescence and early adulthood. We conclude that Respondents were indeed somewhat less stable workers as adolescents than as young adults. Two major explanations were identified: 1. Adolescents are more likely to be students, single, and childless. These characteristics are, as one would expect, conducive to job-instability. 2. The most prevalent reason for leaving a job among adolescents is conscription into the military. Age group differences in job characteristics, human capital, and cognitive ability were found to affect stability in ways very similar to those revealed by previous studies (Sorensen and Tuma, 1981; Carroll and Mayer, 1985), but were not found to explain age variations in stability. When all of these sets of factors where controlled, Respondents appear to have been more stable workers as adolescents than as young adults. They were less likely to shift from one job to the next at any point in time, irrespective of whether the move involved upward, downward,

or lateral mobility along the occupational prestige hierarchy. They were also less likely to shift from employment to unemployment. Thus these findings are clearly at odds with a model which assumes that with age (at least in early adulthood), workers' careers become more orderly. We find no indication that in Israel the transition into adulthood involves either stabilization in jobs or that job shifts are more likely to be in an upward direction.

How do we account for the inherent stability of teenage workers in Israel? At this point we can at least speculate with respect to voluntary job termination: most adolescent jobs begin between ages sixteen and eighteen (Figure 1), that is, within two years or less of conscription. These jobs are probably viewed by their incumbents as temporary because the date of induction is pre-set to age eighteen or so. On the one hand, such jobs may be discarded more readily. On the other hand, why bother leaving a job when any subsequent job will surely be very brief? Thus, adolescents simply "wait out time" in their jobs. A similar understanding of the temporary nature of the jobs entered at adolescence may prevail on the part of employers.

The search for career lines and experimentation with different types of jobs begins in the post-military ages. This is reflected in the higher rates of movement in the 21-27 age category (Table 4). We expect that in later ages the pattern reverses into increasing stability, but the inflection point is outside the age range for which our data are available. In unreported analyses we distinguish

between four age categories (14-17; 18-20; 21-24; 25-27) and find that instability rises monotonically throughout the observed age range.

An additional important short-coming of our work history data is that they pertain to a historical period (1968-1980) of full employment in Israel. In periods of full employment job shifts are more likely to be initiated by employees rather than by employers. Thus our findings are probably applicable primarily to voluntary job instability as it relates to age in the early life course of Israeli men. At the present-day unemployment rates in Israel, job shifts are probably more often initiated by employers. Age may have a very different relationship to involuntary job terminations than our data and analysis suggest.

References

Allison, Paul

  1984   <u>Event History Analysis: Regression for Longitudinal Event</u>
           <u>Data</u>. Beverly Hills: Sage Publications.

Berg, Ivar

  1970   <u>Education and Jobs: The Great Training Robbery</u>. New York:
           Praeger.

Bills, David B.

  in-   "Educational credentials and hiring decisions: What employers

  progress look for in entry level employees." Department of Sociology,
           University of Iowa.

Blau, Peter and Otis D. Duncan

  1967   <u>The American Occupational Structure</u>. New York: Wiley .

Brown, Charles

  1982   "Dead-end jobs and youth unemployment." Pp. 427-451 in
           Richard B. Freeman and David A. Wise (eds.) <u>The Youth</u>
           <u>Labor Market Problem: Its Nature, Causes, and</u>
           <u>Consequences</u>. Chicago: The University of Chicago Press.

Carroll, Glen R. and Karl U. Mayer

  1985   "Job-shift patterns in the Federal Republic of Germany: The
           effects of social class, industrial sector, and organizational
           size." Max-Planck-Institut fuer Bildungsforschung, Berlin.

Coleman, James, S.

  1984   "The transition from school to work."
           <u>Research in Social</u>
           <u>Stratification and Mobility</u>. Vol. 3: 27-59.

Cox, D.R.

1972   "Regression models and life tables." Journal of the Royal
        Statistical Society, Series B 34: 187-202.

Doeringer, P., and M.J. Piore

1971   Internal Labor Markets and Manpower Analysis. Lexington,
        Massachusetts: Heath.

Erikson, Erik

1968   Identity Youth and Crisis. New York: Norton.

Featherman, David L. and Robert M. Hauser

1978   Opportunity and Change. New York: Academic Press.

Hachen, David

1983   "Class and career mobility." Department of Sociology
        Working Paper Number 12, University of Wisconsin-Madison.

Harrell, Frank

1980   "The PHGLM Procedure" Pp. 119-131 in SAS Supplemental
        Library User's Guide 1980 Edition Cary, North
        Carolina: SAS Institute Inc.

Hauser, Philip, M.,

1979   "Introduction and overview" in P.M. Hauser (ed.) World
        Population and Development: Challenges and Prospects.
        Syracuse: Syracuse University Press.

Hogan, Dennis P.

1982   Transitions and Social Change: The Early Lives of American
        Men. New York: Academic Press.

Kahane, Reuven and Laura Starr

1984   Education and Work: Processes of Vocational
        Socialization in Israel. Jerusalem: Magnes.

Kallenberg, Arne L., and Aage B. Sorensen

1979   "The sociology of labor markets." Annual Review of
        Sociology 5:351-79.

Kraus, Vered

1976   "Social grading of occupations in Israel." Unpublished
        Ph.D. dissertation. Jerusalem: The Hebrew University.

Mare, Robert D.

1981   "Change and stability in educational stratification."
        American Sociological Review 46:72-88.

Mare, Robert, D., Christopher Winship, and Warren N. Kubitschek

1984   "The transition from youth to adulthood: Understanding the
        age pattern of employment." American Journal of Sociology
        90:2 326-58.

Matras, Judah

1984   "On schooling and employment in the transition of Israeli
        males to adulthood." Unpublished manuscript, Max-Planck
        Institute fur Bildungsforschung, Berlin.

Matras, Judah, Gila Noam, and Itzhak Bar-Haim

1984   Young Israelis on the threshold: A study of the 1954 cohort
        of Israeli men." Jerusalem: Brookdale Institute Discussion
        Paper D0284.

Meyer, Robert H., and David A. Wise

1982   "High school preparation and early labor force experience."
        Pp. 277-344 in Richard E. Freeman and David A. Wise (eds.)
        The Youth Labor Market Problem: Its Nature, Causes, and
        Consequences. Chicago: The University of Chicago Press.

Mincer, J.

1974   Schooling, Experience and Earnings. New York: National
        Bureau of Economic Research.

Nahon, Yaacov

forth-   "The educational location of Criental Jews in Israel."

coming   Jerusalem: Jerusalem Institute for Israel Studies.

Csterman, Paul,

1980   Getting Started: The Youth Labor Market. Cambridge, Mass.:

           MIT Press.

Raven,J.C.

1958   Standard Progressive Matrices. London: H.K. Lewis an   d Co.

Shavit-Streifler, Yossi, J.

1983   "Tracking in Israeli education: Its consequences for ethnic

           inequality in educational attainment." Unpublished Ph.D.

           dissertation. Madison, Wisconsin: University of Wisconsin-

           Madison.

Shavit, Yossi

1984   "Tracking and ethnicity in Israeli secondary education."

           American Sociological Review 49: 210-20.

Shavit, Yossi and David L. Featherman

1985   "Constancy and change in teenage intelligence: The effects

           of secondary education in Israel." Paper to be presented at

           the Annual Meeting of the American Sociological Association,

           Washington, D.C.

Shavit, Yossi and Arye Rattner

1986   "Age, crime, and the early life course in Israel." Paper

           presented at the Annual Meeting of the American

           Sociological Association, New York.

Sorensen, Aage B. and Nancy E. Tuma

1981   "Labor mobility structures and job mobility." Research in

           Social Stratification and Mobility. Vol. 1 67-94.

Smooha, Sammy

  1978   Israel: Pluralism and Conflict. Berkeley: University of

           California Press.

Spenner, Kenneth I., Luther B. Otto, and Vaugh R. A. Call

  1982   Careers Lines and Careers. Lexington, Mass: Lexington

           Books.

Spilerman, Seymour

  1977   "Careers, labor market structure and socioeconomic

           achievement." American Journal of Sociology 83:551-93.

Thurow, Lester C.

  1975   Generating Inequality: Mechanisms of Distribution in the

           U.S. Economy. New York: Basic Books, Inc.

Tuma, Nancy, E. and Michael T. Hannan

  1984   Social Dynamics: Models and Methods. New York: Academic

           Press, Inc.

Table 1: Means and standard deviations of variables by age groups.

Variable	age groups 14-17		age groups 18-20		age groups 21+	
	mean	S.D.	mean	S.D.	mean	S.D.
Job characteristics						
PUBLIC (public sector)	0.02	0.13	0.13	0.34	0.09	0.29
PRESTIGE (prestige of occupation)	19.98	13.27	24.30	16.68	31.81	22.08
NSUBS (number of subordinates)	0.41	1.66	1.50	5.47	2.13	8.07
HOURS (number of hours worked per week)	42.72	10.89	46.34	13.03	46.13	13.86
Worker's human capital						
SECONDARY (some secondary education)	0.76	0.43	0.84	0.36	0.92	0.27
MATRICULATION (high school diploma)	0.04	0.20	0.17	0.38	0.27	0.44
UNIVERSITY (some higher education)	0.01	0.07	0.06	0.23	0.24	0.42
DEGREE (a university degree)	0.00	0.04	0.02	0.13	0.13	0.34
EXPERIENCE (months of labor force experience)	5.16	9.72	17.77	20.17	32.63	28.72
LIFE COURSE VARIABLES						
STUDY (enrolled in school at end of job spell)	0.35	0.48	0.10	0.30	0.14	0.35
MARRIED (married at end of job spell)	0.02	0.13	0.24	0.43	0.46	0.50
CHILD (parent at end of job spell)	0.01	0.11	0.17	0.38	0.29	0.45
IQ (verbal and analytic intelligence)	51.65	18.20	54.97	18.86	57.63	17.72
Social background						
POPSEI (father's occupational prestige)	33.37	13.43	37.27	18.88	35.49	16.51
ETHNIC (Sephardi origins)	0.82	0.38	0.75	0.43	0.75	0.43
DURATION (number of months worked on the job)	12.03	11.94	13.00	12.80	13.85	12.30
Censoring by interview	0.01	0.11	0.09	0.28	0.35	0.48
Sequential number of spell in work history	1.71	1.06	2.59	1.60	3.77	2.22
Number of jobs	602		247		1599	

Table 2: Proportional hazards models of job leaving: all non-self
employed jobs entered since age fourteen (censoring defined by
time of interview).

Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
AGES 18-20	-0.089	-0.117	-0.182*	-0.109	-0.188*	-0.253*	0.013	-0.152	0.008	0.089
AGES 21-27	-0.511*	-0.550*	-0.595*	-0.526*	-0.464*	-0.674*	-0.290*	-0.402*	-0.280*	-0.065
POPSEI			0.003*	0.005*	0.002	0.002	0.001	0.002	0.001	0.001
ETHNIC			0.081	0.094	0.047	0.008*	0.124	0.110	0.118	0.132
PRESTIGE				-0.003*				-0.005*	-0.005*	-0.004*
PUBLIC				-0.505*				-0.641*	-0.606*	-0.510*
NSUBS				-0.024*				-0.028*	-0.026*	-0.023*
HOURS				0.002				0.004	0.007*	0.007
SECONDAR					0.045			-0.099	-0.136	-0.177*
MATRICULATION					0.264*			0.006	-0.145	-0.215
UNIVERSITY					-0.326*			-0.420*	-0.691*	-0.732*
DEGREE					-0.319*			-0.512*	-0.877*	-0.951*
EXPERIENCE					-0.012*			-0.013*	-0.012*	-0.010*
IQ						0.008*		0.010*	0.009*	0.008*
STUDY							0.563*		0.806*	0.765*
MARRIED							-0.380*			-0.380*
CHILD							-0.620*			-0.536*
SEQ1		-0.446*	-0.337	-0.250	-0.961*	-0.365	-0.508*	-1.003*	-1.013*	-1.028*
SEQ2		-0.381	-0.263	-0.215	-0.708*	-0.278	-0.389	-0.787*	-0.791*	-0.824*
SEQ3		-0.338	-0.203	-0.161	-0.526*	-0.171	-0.276	-0.551*	-0.585*	-0.586*
SEQ4		-0.377	-0.211	-0.120	-0.460	-0.212	-0.335	-0.463	-0.497*	-0.533*
SEQ5		-0.437	-0.255	-0.209	-0.432	-0.252	-0.388	-0.486	-0.561*	-0.552*
SEQ6		-0.590*	-0.422	-0.386	-0.574*	-0.443	-0.418	-0.660*	-0.675*	-0.608
SEQ7		-0.275	-0.201	-0.147	-0.275	-0.231	-0.246	-0.321	-0.387	-0.373
SEQ8		-0.408	-0.275	-0.131	-0.385	-0.232	-0.463	-0.228	-0.384	-0.457
SEQ9		-0.290	-0.476	-0.507	0.374	0.403	0.219	0.344	0.196	0.135
Chi-Square	103.12	118.14	110.21	147.20	194.96	131.59	353.86	262.82	372.62	483.77

* Parameter at least twice its standard error.

Table 3: Percent distribution of job spells
by age and destination.

Destination	Age Groups 14-17	18-20	21+
upward move	5.7	8.9	12.3
downward move	3.3	13.0	9.4
lateral move	15.8	19.4	22.5
military	33.7	17.4	0.6
unemployment*	15.3	27.1	14.0
school	25.1	5.7	6.0
censored by interview	1.1	8.5	35.2
total	100.0	100.0	100.0

* Unemployment is defined as a spell during which
  respondent was neither in school, nor in the
  military, nor employed in a job.

Table 4: Proportional hazards models of job leaving by destination of move (all models control for sequential position of job in the work history).

| Independent Variables | censored by int'view & military (1) | moved to another job | | | | unemployment# (6) |
		all (2)	upward (3)	downward (4)	lateral (5)	
AGES 18-20	0.381*	0.728*	0.704*	1.172*	0.433	0.575*
AGES 21-27	0.477*	1.038*	1.369*	1.295*	0.847*	0.346*
POPSEI	0.001	0.002	0.008	0.003	-0.002	-0.000
ETHNIC	0.133	0.097	-0.020	-0.091	0.268	0.084
PRESTIGE	-0.004*	-0.004	-0.043*	0.038*	-0.009*	-0.001
PUBLIC	-0.506*	-0.440*	-0.557	-0.111	-0.459	-0.732*
NSUBS	-0.022*	-0.020*	-0.020	-0.054	-0.011	-0.061*
HOURS	0.008*	0.001	0.007	-0.001	0.000	0.010*
SECONDAR	-0.392*	-0.130	0.091	-0.205	-0.187	-0.432*
MATRICULATION	-0.620*	-0.351	0.848*	-1.273*	-0.871*	-0.280*
UNIVERSITY	-1.067*	-0.466	0.131	-1.696*	-0.320	-0.802*
DEGREE	-1.381*	-0.905*	0.131	-1.806*	-1.402*	-0.868*
EXPERIENCE	-0.011*	-0.010*	-0.014*	-0.009	-0.007*	-0.009*
IQ	0.009*	0.001	0.003	-0.007	0.001	0.004
STUDY	1.008*	0.133	0.086	0.019	0.221	-
MARRIED	-0.323*	-0.227*	-0.220	-0.561*	-0.115	-0.838*
CHILD	-0.509*	-0.574*	-0.287	-0.672*	-0.713*	-0.453
Chi-Square	422.15	198.98	142.36	140.34	132.26	207.96

* Parameter at least twice its standard error.
# Unemployment is defined as being neither in school, nor in the military, not employed.
- No vairance in independent variable for uncensored cases.

FIGURE 1: PERCENT OF THE COHORT WHO WERE EMPLOYED, BY ETHNICITY AND AGE.

ETHNICITY —— As-Af —— Eu-Am
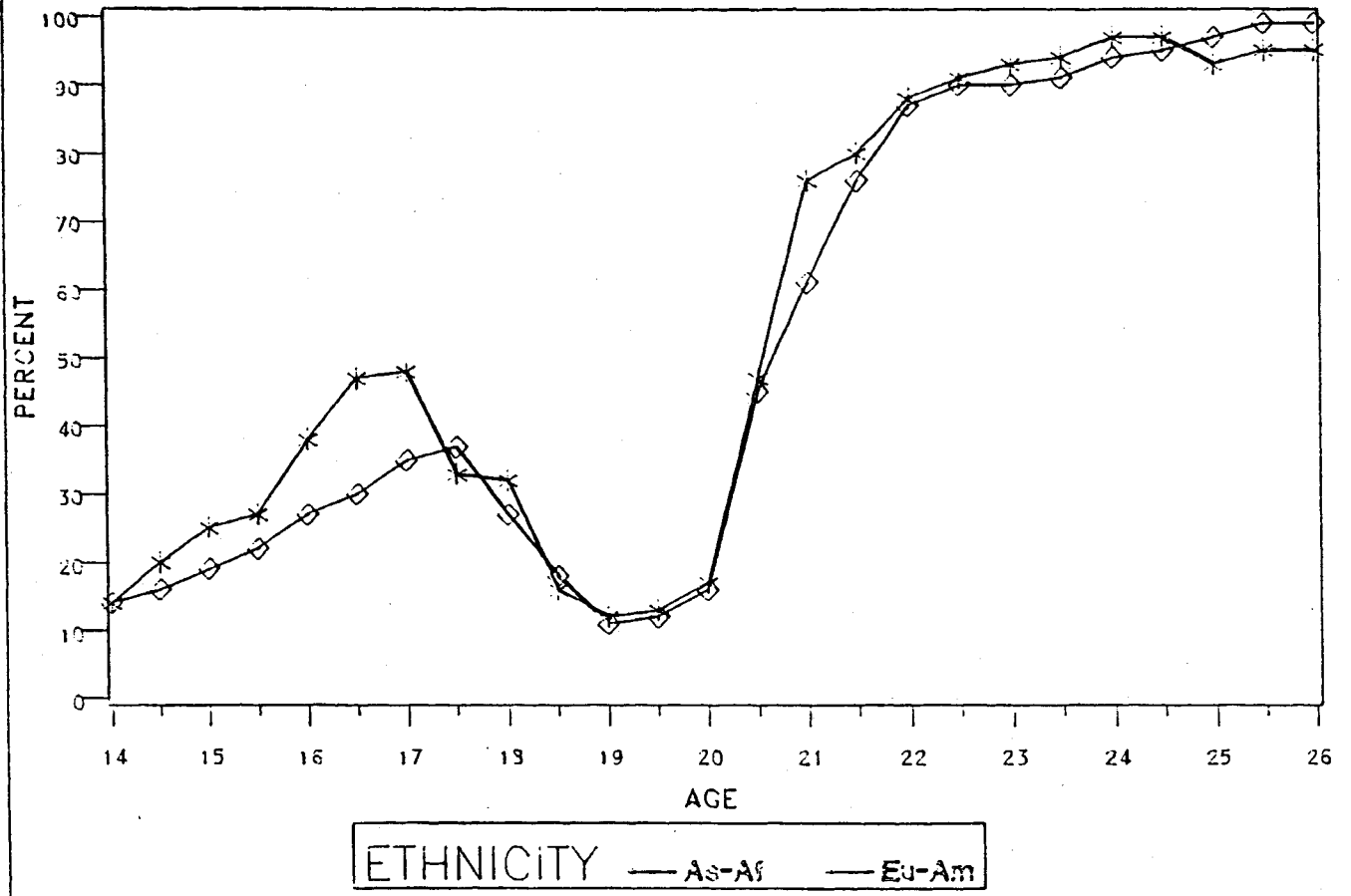
Figure 2: MEAN NUMBER OF HOURS WORKED PER WEEK BY ETHNICITY AND AGE.
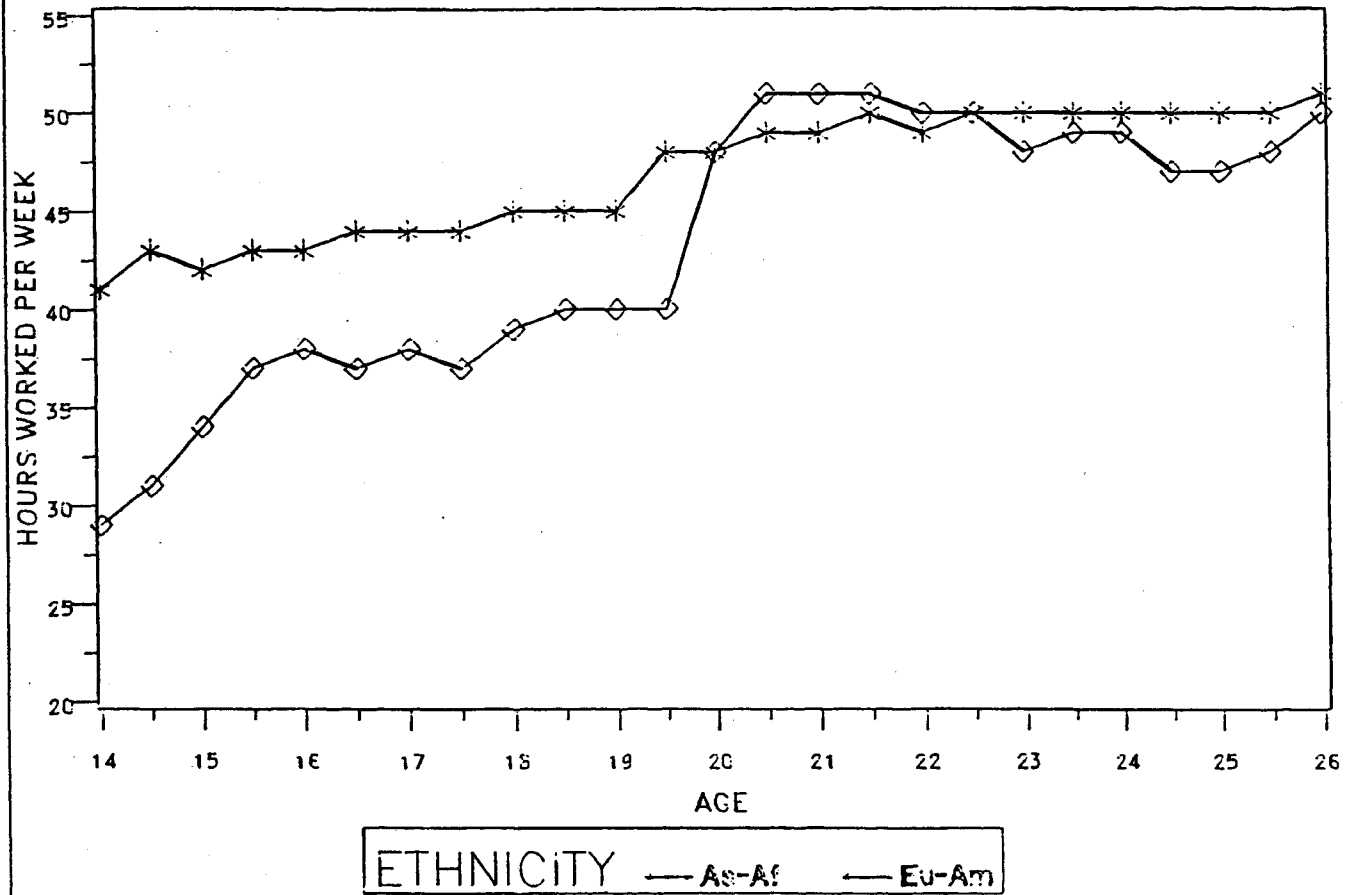
Figure 3: CUMULATIVE PROPORTIONS OF JOBS SURVIVING BY DURATION AND AGE GROUPS.
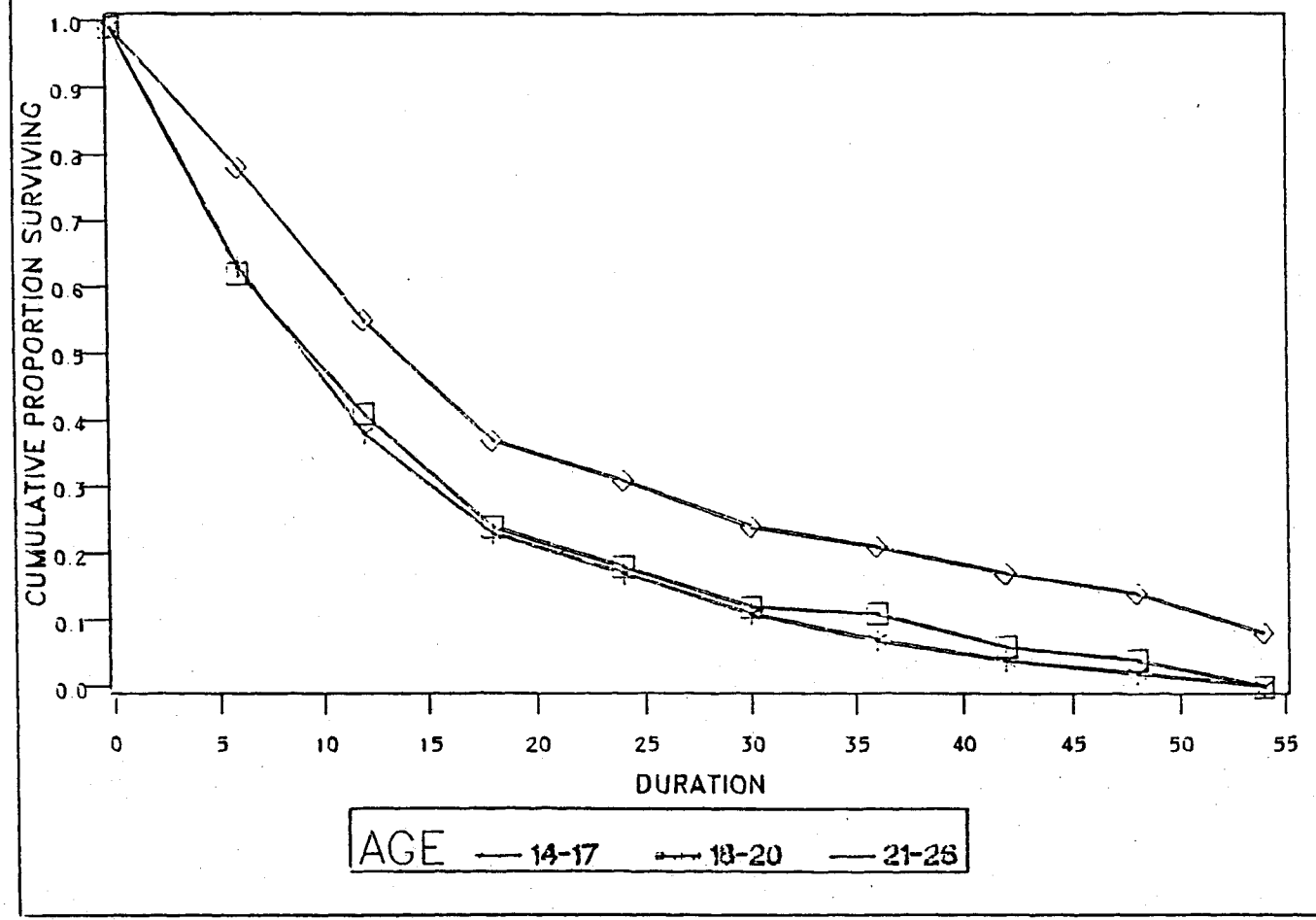
Jobs and classes: Structural constraints on career mobility

Karl Ulrich Mayer and Glenn R. Carroll

INTRODUCTION

Sociological theory offers two fundamentally different views of the relationship between class and mobility. First, class structure is frequently taken to be the result of both collective and individual social mobility. According to this position, high rates of mobility contribute to the transformation or even dissolution of a given class structure. Such reasoning can be found in Marx' writings in his discussions of the "declassement" of small proprietors into the working class and of the coalition of bourgeois intellectuals with the working class. In both instances, Marx uses mobility as a key component of the process of class formation. Pushing this line of thinking to the extreme Schelsky (1965,1968) has claimed that high mobility rates in post-war Germany have dissolved the traditional class structure of this country altogether.

The second and opposite view holds that mobility has nothing to

do with the actual constitution of class structure. Instead,
mobility is seen to affect the consciousness, internal coherence
and potential for collective action of a specific class.
Dahrendorf (1969:108/109), for example, writes that, "social
classes ... are phenomena which at least potentially exist
independent of the mode of recruitment and rate of fluctuation of
their members ... in this sense social mobility as such is
irrelevant to the problem of the existence of classes." The
argument thus closely resembles Schumpeter's (1953:171) famous
analogy: "...each class is like a hotel or an omnibus for the
duration of its collective existence, which is always occupied,
but always by other persons." This view, like the first, has
also been defended and used by Marxist theorists.

In fact, much of the recent discussion about class makes use of
various shades of the age-old debate about the relevance or
irrelevance of mobility [e.g. Wright,1979 (mobility is
irrelevant); Goldthorpe,1980 (class structure is not constituted
by mobility but mobility is important for class coherence and
organization); Giddens, 1973 (mobility is important for the
transformation of economic categories into social
collectivities)]. Typically, the debate is conducted with
reference to empirical research on inter-generational
occupational mobility (Goldthorpe, 1980; Mayer,1977) . The class
discontinuities between fathers and sons, reflecting primarily
massive changes in the occupational structure over time, may
indeed be consequential for class formation (e.g. if most blue
collar workers are sons of self-employed farmers). However,
intragenerational, worklife or career mobility surely must have a
stronger impact on class formation. Fluctuations across classes
in the course of a career should undermine class loyalty as well
as loosen any potential homogeneity in the material conditions or
orientations of a class. Conversely, if class boundaries can be
easily transcended during a worklife, then the salience of
assumed class distinctions might well be questioned.

In industrial societies, career class mobility takes the form of
job shifts. Yet the relationship between class mobility and job
mobility remains a sociological mystery. Conceptually, it is

clear that every class change involves a job shift but that many
job shifts do not involve a class change. Empirically,
intergenerational class mobility, career class mobility, and job
shifts have each been studied thoroughly but in isolation. As a
result, we still do not know the answers to such basic questions
as : How do classes differ with respect to job-shift patterns?
Which classes protect members from the labor market and which
expose them? Which classes offer upward mobility passages? How
many job shifts does it take to get to a given class?

There is also a diachronic dimension to class entities which has
been overlooked by theorists and which identifies structures with
cross-sectional patterns (Mayer,1986). Marx clearly saw this
dimension, for it was essential to him that classes be
self-reproductive over time. According to Schumpeter's omnibus
imagery, interest should be focused not only on which buses are
boarded but also on where the buses are going. The class
experience of a collectivity can be directly inferred from the
social structural path over which this collectivity travels
during the worklife years. Thus, the class condition of a
blue-collar worker includes his declining market value as a
result of his failing physical strength, and his pension rights
which are precisely fixed on the basis of the duration of his
class membership.

Intra-career shifts and conventional mobility analysis

A great number of empirical studies on social class involve the
analysis of inter- or intragenerational mobility tables. With few
exceptions, the data in mobility tables are cross-sectional in
that they record the respective class of fathers and sons at one
specific point in time. For intergenerational mobility it has
become conventional to take age 14 or 15 of the respondent as
reference time for recording the father's occupation or class,
while the son's (respondent's) class is recorded at time of the
interview. For intragenerational mobility, the first job is used
as a reference for the origin position.

To the extent that the cross-section consists of a probability sample of either fathers or sons, the relevant marginal distribution of classes represents a good approximation of the societal class structure (Mayer 1979a). Further, to the extent that individuals do not change classes across their life, comparisons of the marginal distributions of the sons against the fathers provides information about the change in societal class structure across two generations as well as the pattern of intergenerational transmission of class.

Suppose, however, that the typical life course involves some "bouncing around" early in the career - often with marginal activities like farm labor in the first job - such that it is not until the third or fourth job around 25 or 30 that a son settles into a relatively permanent social class (Blossfeld 1986a). Under such conditions, a typical mobility table is likely to lead to erroneous inferences to the extent that sons under age 25 or 30 are represented in the sample since their class is likely to be unidentified, or at least incorrectly assumed to be stable. The irony here is that if no sons under age 25 or 30 are present in the data, the table may be acceptable for studying intergenerational mobility but it will be of little use for studying societal class structure since its representativeness is highly implausible (unless of course, the societal demographic structure is equally perverse). Similarly, we have to expect that occupational degradation occurs to a considerable extent in the last phase of worklife. Such degradation would affect both fathers and sons in a mobility table (Mayer/Müller 1972).

Thus, we have to admit the possibility that the typical life course involves several changes in class across the entire career. That is, let the probability of sons changing class at any moment in time be a finite, non-negligible number. Let the same be true for fathers. Clearly, under these conditions, a cross-sectional snapshot of the respective class of fathers and sons will be highly misleading - if not entirely fallacious - as it concerns intergenerational mobility (although of course, it may still inform perfectly as to the societal class structure).

From these hypothetical scenarios, it is obvious that a finding
of regular class changing may greatly undermine the empirical
research based on mobility tables. Yet, despite the importance of
the assumption of individual class constancy, little is in fact
known about such phenomena. At any rate, the available
descriptive studies are restricted in the sense that they look at
transitions only from first class to class at the time of the
interview (Goldthorpe,1980), or only between a given span of
years, say 1965 to 1970 (König/Müller,1986). Although break-downs
according to age groups or birth cohorts may guard against
certain erroneous inferences, they do not alleviate the basic
flaw. What is sorely needed here is systematic evidence of the
extent of class changes, something that can best be accomplished
by studying the working lives of individuals.

Studying class changes among individuals has other complications.
First,as we stated above, in modern economies classes are
intimately related to jobs. Yet this does not mean that a study
of class changes amounts to a study of job shifts - one must
change jobs to change class but one can change jobs without
changing class. Second, job changes and class changes are
constrained by other structural factors, most notably,
organizational and industrial barriers. These must be
disentangled if one is to sort out the true dynamics of social
class. Third, the educational system plays a crucial role in
affecting and reproducing both jobs and classes. This implies
that not only career histories but life histories must be
collected and examined.

Although several studies have examined the relationship of job
mobility and class mobility across a career (Breiger,1981;
Goldthorpe,1980; Haller/ Hodge,1981; König/Müller,1986;
Snipp,1985), none has dealt with the full complexity of the
process. Perhaps the most comprehensive treatment was our earlier
paper (Carroll and Mayer, 1986) in which we showed the strong
complementary effects of social class, industrial sector and
organizational size on job-shift patterns.

Yet, this treatment was also deficient when viewed from a class

perspective. There are still no answers to the following questions: under which conditions does class membership make a difference? how does social class affect the life chances of individuals ? have the class experiences of individuals changed across time ?

In this paper we address these questions directly using life history data from the Federal Republic of Germany. In the next section, we elaborate the conception of class which forms the frame of reference for the empirical work and discuss theories of class mobility. Within this section we also review our previous work and discuss the theoretical questions raised therein and in research of others. We then discuss the data and models we use in the analysis. The fourth section of the paper presents our empirical findings. In the final section we discuss the broader theoretical and empirical implications of our research.

CONCEPTIONS OF CLASS AND THEORIES OF CLASS MOBILITY

Distinguishing classes

Any analysis of class mobility presupposes a conceptual scheme and measurement rules for mapping jobs into classes. Two ways of conceptualizing classes may be distinguished. The first we call structuralist, since the class structure is deduced from theoretical reasoning. An example would be Wright's(1979) class scheme where he supplements Marx' two polarized classes by "contradictory class locations" like managers. We applied this scheme in our earlier study on job shifts (Carroll/Mayer 1986).[1] Another example would be Dahrendorf's(1969) incorporation of the authority dimension or Roemer's(1982) proposals for identifying exploited classes. Classes are here initially defined independently of individual and collective behavior and constructs like "structural location within the social relations of production" and "latent interests" are used to this end.

The second way of conceptualizing classes may be traced back to

Weber where the criteria of class are clearly specified: classes
exist to the extent that groups share a common market condition
as the decisive basis for their specific life chances (Weber,
1964:679,680). However, it is taken to be an empirically open
question which particular classes form in a given society. As in
Weber's famous definition of a social class mobility itself might
be used as a criterion to determine class boundaries. Following
this direction one would attempt to establish class categories on
the basis of the mobility patterns between a larger number of
occupational categories (Breiger,1981:580).

We will follow here a position where classes and their boundaries
are defined independently of the mobility process on the basis of
homogeneous market and working conditions. The intention is to
combine into a social class "occupations whose incumbents will
typically share in broadly similar market and work situations
...,on the one hand, in terms of their sources and levels of
income, their degree of economic security and chances of economic
advancement; and, on the other, in their location within the
systems of authority and control governing the process of
production in which they are engaged, and hence in their degree
of autonomy in performing their work-tasks and roles."
(Goldthorpe, 1980:39).

This allows,in a next step,to ask about typical patterns of
individual and collective class mobility and to relate
differential mobility patterns to the criteria which served to
distinguish classes: ".. one may think of class positions having
also inherent mobility propensities which will themselves exert
an influence ... independently of that exerted by the relative
sizes of classes" (Goldthorpe, 1980:39).

The class scheme which Goldthorpe has developed shows two
advantages in addition to this clear theoretical rationale. It
has precisely specified measurement rules to map occupations and
jobs into classes which make use of the information on both
occupational activity and employment status. Further, it has come
to be widely used in cross-national research (Erikson/Goldthorpe

1985, König/Müller 1986, Featherman/Selbee 1986) and thus allows to relate our findings to a growing body of literature.

Goldthorpe's classes and a brief description of their underlying occupations are as follows. (All descriptions are excerpted from Goldthorpe, 1980:39-41):

CLASS I: All higher grade professionals, self-employed or salaried; higher-grade administrators and officials in central and local government and in public and private enterprises (including company directors); managers in large industrial establishments; and large proprietors.

CLASS II: Lower-grade professionals and higher-grade technicians; lower-grade administrators and officials; managers in small business and industrial establishments and in services; supervisors of non-manual employees.

CLASS III: Routine non-manual - largely clerical - employees in administration and commerce; sales personnel; and other rank-and-file employees in services.

CLASS IV: Small proprietors including farmers and small-holders; all other 'own account' workers apart from professionals: a) with employees, b) without employees, c) in agriculture.

CLASS V: Lower-grade technicians whose work is to some extent of a manual character, and supervisors of manual workers.

CLASS VI: Skilled manual wage-workers in all branches of industry, including all who have served apprenticeships and also those who have acquired a relatively high degree of skill through other forms of training.

CLASS VII: a) All manual wage-workers in industry in semi- and unskilled grades; and b) agricultural workers.

Theories of class mobility

There is no systematic body of theory on class mobility over the
life course, not the least because issues of job mobility,
occupational mobility,and class mobility tended to be confused.
We shall take from the available literature and, in part,
formulate ourselves a number of theses which can be tested
empirically.

The "closure" thesis

This view of mobility emphasizes the closed and antagonistic
nature of the class system (Parkin, 1971). Position in the
hierarchy is seen as associated with privilege and opportunity.
Those who stand on top use their control over resources to retain
their position and, when possible, to pass it on to their
children. Because retention of class is usually successful and,
because the hierarchy is a closed system with no additional
higher locations to move in to, mobility from the privileged
classes will be lower than from the less privileged classes. Thus
we expect that there will be an effect of closure due to
advantage and exclusion as well as due to a ceiling effect.

To the extent that classes are hierarchically ordered, distance
effects will operate with the consequence of ordered mobility
rates along the hierarchy. Effects of closure should be much more
visible for intra- than for intergenerational mobility. It should
be easier to keep one's privileges for oneself and one's family
than to pass them on to one or several children. In addition,
intra-career class mobility should show lines of closure more
clearly, since it is much less affected by changes in the
occupational structure.

It is less obvious how closure operates at the bottom of the
class system. There is no way to go further down,i.e. there is a

floor effect. Furthermore, disadvantages should act cumulatively
in a negative manner to suppress upward mobility - as a cycle of
deprivation. Both conditions should decrease mobility. Also,
there may be what Parkin (1974) called 'solidaristic
exclusion',i.e. a social exclusivity of the underdogs where lack
of individual power and economic resources is compensated by
mutual support and collective representation or where the pride
of the working class or craft group leads to immobility even when
advantageous opportunities are available outside. The question,
then, is where such solidaristic exclusion, craft and trade
traditions, and other kinds of collective identification will
show itself. It seems much more likely to be the case in the
class of skilled workers or even production supervisors than in
the class VII of unskilled workers.

The structural dominance thesis

Recent research on social stratification has highlighted social
structural explanations of attainment, as opposed to earlier
individualistic explanations. Three social structures have been
argued as playing the dominant role in career outcomes:
organizations (Baron and Bielby 1980;1984), industry segments
(Averitt 1968; Stinchcombe 1979), and social class (Wright,
1979). In an earlier paper (Carroll and Mayer, 1986), we
demonstrated that the effects of these three structures were
complementary but not equal. At some times, organizational
structures would dominate the mobility process, while at other
times, industry segment and social class would dominate.

Internal labor markets provide a useful starting point from which
to untangle this complex relationship. As is well known, firms
with internal labor markets promote employees regularly and often
in lock-step, on rationalized bureaucratic career ladders.
Mobility within the firm can be predicted fairly well simply on
the basis of time in the current salary grade (see, e.g.,
Petersen and Spilerman, 1986). By a Weberian view of social
class, the employees of a firm with an internal labor market may
be said to compose a single class to the extent that their

mobility experiences are homogeneous (which , however, may not be true for entry level and certain higher level-exempt employees). Indeed, a reasonable interpretation of the literature on internal labor markets is that it sees employees working within such a firm as having qualitatively different career chances than others - they are, so to speak, a class apart.

But the distinctiveness of an internal labor market does not arise from a simple enhancement of working conditions, job security and promotion opportunities. Along with these factors comes a conflation of the usual social class differences. If it is to be taken as legitimate, the career ladder of an internal labor market must be rationalized in such a way that it is perceived as fair, just and egalitarian. Consequently, rules are constructed to prevent moving too many levels too quickly. Likewise, personnel procedures for different occupations become standardized, often by providing manual workers with better than usual conditions of work.

Where, then, will social class be the dominant structure shaping career mobility? The key to the answer lies in the conceptualization of class itself. Marx' original formulation rested squarely on the analysis of the social relations of production as it concerned the exploitation of labor. Nowhere was this more visible than in the thriving industry of his time, and that which he and Engels studied at length, textile manufacturing.

The class framework we use here emphasizes both authority relations and market circumstances. Neither should be a particularly strong factor in internal labor markets (with often considerable union power) or in industrial sectors where authority relations are blurred by professionalism and other status characteristics. Instead, we state the proposition that social class structures mobility in markets and in industrial sectors where labor relations are exploitative and where property rights are weak. A somewhat different argument was put forth by Max Weber in his discussion of bureaucracy (1978:225/226). He argued that bureaucratic organization leads to social levelling

in recruitment since technical competence is being used as the
sole selection criterion.

The life-course thesis

Sociological research on life course examines the ordered
sequences of events in the lives of individuals and how these
change over historical time. Class mobility in a career may be
not only frequent but also structured systematically in
correspondence with age or labor force experience.

There is clear evidence that men do change classes across their
careers. In a recent study, König and Müller (1986) compared the
career mobility of men in West Germany and France between 1965
and 1970. In Germany, 13 % were mobile between the classes
(defined according to the Goldthorpe class scheme), in France
21%. The authors find a clear age-dependency. Looking at the
15-year brackets from age 20-34,35-49 and 50-64, the proportion
of men mobile between classes declines in Germany from 12 to 9 to
5%, and in France from 30 to 14 to 8%. In comparing the first and
current occupations of a large sample of American men, Snipp
(1985), for instance, finds that over 23 % of those with a manual
first job shifted into a later non-manual job. Shifts in the
other direction were almost as prevalent. Similarly, Goldthorpe's
(1980:51-2) data on first class based on first full-time jobs
shows striking differences to class held in 1972 for men aged 35
and over. Over 65% of these men started out in one of the two
lower classes; by 1972 fewer than 43% remained in these classes.
In the other direction, classes I and II showed remarkable
increases. Only 9 % had a first job in either of these two
classes, but by age 35 over 27% did.

Is there any age-related pattern to this jumping from class to
class? Two bodies of theories suggest there is. The first has
been around in mobility literature for some time and is
occasionally referred to as the "counter-mobility thesis" (see
Girod, 1971; Bertaux,1974; Bernard and Renaud, 1976). It argues
that career mobility eventually moves one back to the class of
his or her parents, but only after some initial shifting around,

usually of a downward character. Offspring of upper-class
parents, in particular, are seen as likely to move downward early
in their careers and then to advance generally back to the upper
classes by middle age.

A second body of relevant thought comes from the job search
literature (see March and March, 1981). Many of the models found
here also suggest a lot of early shifting between classes but not
necessarily in the direction suggested by the counter-mobility
theorists. Instead, this literature stresses the stability
arising from good person-job matches, which often require time
and experience to obtain. The intervening period is characterized
by the instability created by poor person-job matches. Both the
counter-mobility thesis and the job search model stress the roles
of individual action and career achievement.

Higher rates of early class mobility could also be the result of
class boundaries which surely are less capable of discriminating
young persons than older ones. Cultural, linguistic, educational
and property attributes - the material from which social barriers
are made - are more equally distributed amoung youth. They are
also more easily disguised and imitated at the younger ages. All
this serves to make social classes much more permeable for the
young than for the middle-aged or elderly.

In contrast, we also have to consider the possibility that age
groups as such constitute specific labor market segments where
both very young and very old workers are statistically
discriminated against and tend to be confined to marginal
employment. We also know from the economic literature
(Oppenheimer, 1974) that young persons can afford to enter
employment where substantive interests are more important than
the calculation of income returns. The "life-cycle-squeeze", with
its constraints in providing economically for a growing family
without a fully employed wife, may force people to change jobs
with the consequence of a change in class position.

Thus - according to the life-course thesis - one would
hypothesize that social class mobility is much more likely in

the early working life.[2]

The rationalization thesis

Work and labor relations have changed profoundly across history and many social theorists believe these deep transformations are continuing today in an especially intensified manner. Two general arguments have been advanced about the nature of post-industrial changes in the labor force, and, hence, the class structure of modern societies of the Western democratic-capitalist type. The first of these is optimistic and has been argued most forcefully by Bell (1973). Post-industrial society is an information-based one, he argues, and it is characterized by the automation of work, increasing demand of service oriented jobs, and a rise in the number and status of administrative, managerial - in particular - science based professional positions. The overall macroscopic trend is one involving the upgrading, so to speak, of the average status as well as the total distribution of jobs in the entire labor force. Similar predictions can be derived from the longstanding debate on sectoral transformation and, specifically, tertiarization (Müller, 1983).

The second and counter-argument is pessimistic. As put forward by Braverman (1974), it sees modern work as becoming more rationalized in the technical sense of specialization and control. Productivity may increase, especially in the short term, but the cost is an increasingly alienated labor force which becomes further detached from any control over its work and therefore the lives of its members. While possibly being temporarily offset by the growth of the service sector, the overall long-term trend is one involving degradation of work, especially as it concerns specific occupations and industries.

Goldthorpe and Payne (1986) have argued that the two arguments have implications which should be discernible in secular mobility trends. If the workforce is being upgraded, they contend, then the expansion of administrative and managerial positions should be accompanied by enhanced mobility opportunities for individuals

not starting from these classes. Conversely, if the overall trend
is one of degradation, then they claim that upward mobility
should be constricting and that mobility into the manual classes
should be on the rise. They find their England/Wales data do not
fully support the implications of either argument but that it is
in closer agreement with the upgrading trend; the degradation
argument is "flatly contradicted (p.19)". West German data on
intergenerational occupational mobility clearly supports the
upgrading thesis up to 1970 under the assumption that degradation
does not overwhelmingly take place in the set of tasks without a
change in job titles (Mayer, 1979a).

The above observation leads to a more sophisticated version of
the degradation argument. It holds that degradation does not
necessarily imply a shift in the overall distribution of classes,
but rather a transformation in the nature and content of jobs
within a given class. Thus the "expansion of administrative and
managerial positions is more apparent than real. This is so
because many of these positions have either been themselves
degraded into essentially subordinate ones, involving only
routine tasks, or have been created by an upgrading of such
subordinate positions of no more than a nominal or cosmetic kind
." (Goldthorpe and Payne, 1986:20). In the German discussion the
latter position has been put forward by Kosta, Krings,and Lutz
(1970), whereas for many years the major dogma was the
middle-of-the-road position of polarization of work allowing
simultaneously for both upgrading and degrading processes
(Kern/Schumann, 1970).

Once one allows for changes in job content which are not
observable via mobility between jobs, there is no way to arrive
at clear empirical conclusions at the level of the total class
structure. However, one major implication of the rationalization
of work thesis is that classes have become internally
differentiated in a job-related sense. To paraphrase Goldthorpe
(1980:54),"the channels of mobility have changed without
affecting its extent."

Thus we hypothesize that job mobility within classes has
increased steadily in the transition to the post-industrial era.

A closely related idea brings the educational system into play.
Noting that the large mass enrollments of modern educational
systems have increased the "credentialled" population enormously,
Goldthorpe distills from Parkin (1971:62-67) the implication,
that rather than tightening the bond between educational
qualification and social class, the surfeit has actually loosened
(or 'counterbalanced') it. Thus, in addition to a dimunition of
the effects of parent's class over the life time, one may also
expect a lessening and a loosening of the effects of education in
a rationalized society.

The reproduction thesis

Although class structures change - incrementally or in sudden
upheavals - they should be fairly stable from the perspective of
most persons' lifetime. Such a postulated inertial tendency
requires special explanation (Hernes, 1976). There are at least
two insightful directions in which to search for an answer.
First, class structures, like any other social institutions, must
be reproduced on a daily basis. Certain jobs, organizations, and
even "industries"(e.g. child care) and, not least, the state are
devoted to these functions. More significantly, class structures
reproduce themselves on a daily basis because the social
relations of production are necessary conditions for the
existence of wage earners and employers. The conditions for
physical, material, and cultural reproduction are determined by
the market wage (Marx,1970 (1867):181 f.).

The second important aspect of reproduction theory concerns
intergenerational class continuity. The current debate here is
over the intervening mechanisms of transmission, while widespread
class inheritance is often taken for granted. Class may be
reproduced intergenerationally in a direct way through
differential access to family capital, privilege, and social
contacts. Others have been persuaded by Bourdieu
(Bourdieu/Passeron, 1971) that intergenerational reproduction is

more indirect, operating primarily through the educational
system. Despite the apparently egalitarian and meritocratic
nature of modern education, Bourdieu holds that upper class
families get more from it for their children by virtue of higher
investments in cultural capital, which allows fuller exploitation
of curricula, facilities and teachers. In addition, upwardly
mobile families use their newly acquired financial capital to
invest in the educational and cultural capital of their children.

Our interest is with the timing of class reproduction in the life
course, an issue which might clarify the debate over direct and
indirect intergenerational class transmission. By the life course
thesis above, we argued that early class positions were less
stable than those entered later in life. The observation of
counter-mobility, whereby the children of upper class parents
drop in class stature initially and then return to their parents'
class suggests a specific avenue and timing of reproduction. By
looking at the different effects of education and family
background at different points in a career, it seems reasonable
to expect that education forms class memberhip most strongly
early in the career, while family background has stronger and
delayed effects late in a career.

We should point out here that the concept of class reproduction
has been applied to career mobility in a somewhat different
fashion as well. Asendorf- Krings et.al. (1976) and, recently,
Lappe (1985) and Blossfeld(1985) suggest a model of collective
career mobility for the Federal Republic in which class
membership is fixed prior to entry into job trajectories in a
system of apprenticeships and vocational training. The later
class fate of job entrants is highly determined by their initial
level of qualifications. These groups are supposed to be
selective at career entry in regard to social origin, cognitive
ability , schooling and motivation. In addition, it is presumed
that vocational training and early career have powerful
homogenizing effects. The later collective mobility can be termed
a "reproduction trajectory" because the members of such classes
share their specific market conditions diachronically. The

fruitfulness of this concept clearly depends on the degree to
which such collective trajectories can be observed, in contrast
to highly diverging paths, even under the condition of shared
class membership at career entry.

DATA AND MODELS

We use retrospective data from the German Life History Study
directed by Karl Ulrich Mayer (1979c). Three cohorts of men and
women born in the years 1929-31, 1939-41 and 1949-51 were
sampled and interviewed by GETAS, a professional survey research
firm. The sample of 2172 persons is representative for the
Federal Republic and West Berlin and based on information from
2172 persons. Details of the sampling plan, field procedures and
data coding can be found in Brückner et.al.(1984). Descriptive
statistics can be found in Blossfeld/Hamerle/Mayer (1986),
Blossfeld (1985), Mayer/Wagner (1986).

Our confidence in the quality of the data on objective
life-events such as job change, marriage and migration is based
on three pieces of circumstantial evidence: 1) a successful ten
year follow-up pilot study of residential, familial, and
occupational changes of 35 men (Toelke, 1980; Papastefanou,
1980), 2) painstaking checks of the life history protocols which
included a high number of second contacts with the respondents by
interviewers, by mail and telephone, 3) a favorable statistical
comparison of the data with comparable data drawn from the 1971
and 1981 microcensus (Blossfeld, 1986b) and census data on
fertility (Tuma and Huinink, 1986).

Given this evidence, we believe that the retrospective life
history data is at least as good - and probably better - than
the conventional cross-sectional survey. More important than the
absolute measurement quality of the data, however, are questions

of possible systematic recall error that might confound
empirical findings. The most plausible hypotheses of this sort
are:

1) that events closer to the time of the interview will be better
recalled, thus monotonically confounding recall with age or
cohort, and 2) that episodes of longer duration will be
remembered better than those of shorter duration, again causing
interference with age or cohort variables. We cannot, on the
basis of the available data, definitively rule out either
hypothesis. Therefore, we take great care in interpreting the age
and cohort variables in our models. Other variables are
relatively insensitive to these problems.

Our analysis involves primarily the job history component of the
Life History Study. For each respondent, data were collected on
the dates of the beginning and ending of each job held.
Respondents were also asked to identify the occupation, wage
rate, industry and size of firm for each job. Changes in jobs
were flagged when they occured within the same firm.

These basic data were used to code the structural variables of
interest on social class and industrial sector. For the
Goldthorpe (1980) class schema, we used adapatation rules
developed by Walter Müller and his associates at the University
of Mannheim for the German occupational structure.

For industrial sectors, we used the seven-fold schema of
Stinchcombe(1979). Our previous paper describes the rationale and
coding behind this schema in great detail, so we refrain from
discussing it here except to note that the sectors where labor
relations are believed to be most exploitative are labelled
'capitalist' and 'small competitive'. Internal labor markets are
most prevalent in the 'bureaucratic' and 'large-scale
engineering' sectors. Both of these groupings are pertinent to
the structural dominance thesis.(Details on the mapping of
industrial branches into the classification by Stinchcombe can be
found in Appendices A and B.)

Most of the analyses reported below use rate functions to model the effects of variables of interest on job shifts and class shifts. Rate functions specify as the dependent variable the construct

$$r(t) = \lim_{dt \to 0} \frac{Pr(\text{change between } t, t+dt \mid \text{state at } t)}{dt}$$

where the probability is of a job or class change between times t and t+dt, given that a job or class is held at time t (see Tuma and Hannan, 1984, or Blossfeld et al., 1986). To model the effects of independent variables, we use the proportional hazards model of Cox (1972). This model specifies that

$$r(t) = h(t) \exp(b_1 X_1 + \ldots \ldots b_k X_k)$$

where the b coefficients measure the size of the effects of the X exogenous variables; h(t) is an unspecified nuisance function which is assumed to affect each sample unit identically. For estimation, we use partial likelihood techniques (Cox, 1975), which yield unbiased and efficient estimates under reasonable assumptions (Efron, 1977). Use of the estimator requires a "proportionality" assumption that the gradationa in a covariate have constant proportional effects on the rate. Exploratory analyses have shown that for some of our variables, particularly sex, this assumption is violated. Nonetheless, the robustness of these coefficients suggests strongly that single equation models are preferable to those obtained from partitioning the sample.

Other analyses which we report below use the conventional techniques of ordinary least squares and logistic regression. The following is a complete list of the variables we use, along with their assigned names.

Cohort 1939-41: An indicator variable which takes the value of one for respondents born in the second cohort of the sample (between 1939 and 1941 inclusive); otherwise the value is zero.

Cohort 1949-51: An indicator variable which takes the value of one for respondents born in the third cohort of the sample (between 1949 and 1951 inclusive); otherwise the value is zero.

Sex: An indicator variable which takes the value of one for female respondents and the value of zero for men.

Experience: Measured as the number of months elapsed since entry into the first job.

First job: An indicator variable which takes the value of one for the first job held by a respondent; the value is zero for all subsequent jobs.

Job status: Scale of social prestige of occupation. Based on the extensive work of Wegener (1984) using German survey data.

Education: Scale of highest level of general education completed. Takes the values: 1 if Volks-/Hauptschulniveau (elementary school); 2 if Mittlere Reife (middle school degree); 3 if Abitur or Fachabitur (high school degree).

Training: Scale of highest level of occupational education completed. Takes the values: 1 if no vocational training; 2 if apprenticeship; 3 if Meister-/Techniker-degree(technical training); 4 if Fachhochschulniveau (technical college degree); 5 if Universitätsabschluß (university degree).

Lnsize org.: Log of number of employees in organization.

EMPIRICAL ANALYSIS

We shall present the empirical findings in the following
sequence. After an overview of some basic distributions the
impact of class membership on various forms of job mobility is
examined. Then we ask how these impacts of class vary for
different industrial sectors and how mobility between jobs,
sectors and classes are interrelated. In the next steps we model
the individual and collective determinants of any class move in
general and for each origin class separately, the conditions of
initial placement into a social class and of mobility into
particular classes of destination. Finally we try to account for
the number of job shifts within given episodes of class
membership.

Table 1 provides descriptive information on the distribution of
social classes in the life history data. The work histories of
these three cohorts of 2172 West Germans comprise 6732 jobs or
employment episodes. Of these, a majority are in skilled and
unskilled manual classes. Fewer than 5 % are in the higher
professional, administrative, and managerial class.

(Table 1 about here)

Social classes are not equally distributed across industrial
sectors. For illustration, we have chosen three very different
sectors and presented in Table 1 the distribution of jobs in
each. The large-scale engineering sector contains an abundance of
skilled and unskilled manual jobs. Distributionally, the
classical capitalist sector contains even more: over two thirds
of the jobs fall into these classes. By stark contrast, the
professional sector has very few jobs of these kinds -- the bulk
of its jobs are in the higher and lower professional positions.

One way to examine these data is to look at the distribution of
first jobs across the different cohorts. From Table 1 it appears

as though from the period 1945 to 1970 (the time of entry of these cohorts) the class structure has shifted "upward" in its distribution. The professional and upper service classes I and II have increased in abundance while the manual classes, especially the unskilled and semi-skilled class VII, have declined. The self-employment class IV has also steadily dwindled as an initial entry point.[3]

Aggregating these job episodes to class episodes (changes in jobs are ignored unless they occur across two distinct classes) result in 4032 distinct spells for analyses. As Table 1 shows, the mean class episode consists of 1.54 jobs, although for the skilled manual class it goes as high as 1.90. Figure 1 shows the distribution of jobs per class episode. As might be expected, it is highly skewed to the right: The greatest number of class episodes contain only a single job.

(Figure 1 about here)

Job Mobility

Table 2 presents estimates of the effects of social class on various types of job mobility when cohort, sex, education and organizational size are controlled for. The class coefficients are estimates relative to the omitted class, semi- and unskilled manual workers in agriculture (VIIb).

(Table 2 about here)

A number of things about the estimates are noteworthy. First, with the exception of the self-employment class IV, the ordering of the classes in terms of mobility propensities conforms to that expected by the closure thesis. In most equations, the higher professional class I is the least likely of the employed classes to change jobs. The lower service class (II) is next in the ordering; followed then by lower technical and manual supervisory jobs(V); then by routine non-manual jobs (III) and the manual jobs (VI and VII).

Second, consistent with the structural dominance thesis, the effects of social class are significant for the across-firm moves but not for the within-firm moves. By these estimates, internal labor markets seem to smooth over the usual effects of class, perhaps by instituting a different class regime.

Third, the effects of occupational status are overwhelmed by the class variables. In no equation does the status variable achieve statistical significance.[4] This runs counter to a strong interpretation of the closure thesis, such as that advanced by Blau and Duncan (1967). Obviously the class scheme is able to catch quite homogeneous socio-economic groupings, where the internal status differentiation is too small to affect mobility propensities. In contrast, general educational resources remain effective. This implies that school qualifications not only determine access to a given class, but also act as a mobilizing force in job shifts up to a point where skills match the job requirements.

It is noteworthy that occupational training is consequential in addition to class memberhip for three types of job shifts. It is negatively related to downward mobility, mobility across firms and upward mobility across firms. That vocational resources prevent slipping confirms our theoretical expectations. However, that such resources apparently tend to tie workers to their firm seems surprising in the light of theories which postulate an occupational labor market transcending firms (Sengenberger, 1975). It seems rational that firms want to keep the most qualified workers, but obviously they are also able to do so by offering sufficient incentives within the firm. Another contradiction to this theory is provided by the finding that not unskilled workers are most prone to lateral moves across firms but skilled workers.

Fourth, contrary to the life course thesis, first jobs usually show a lower rather than higher rate of movement. This somewhat unusual finding is perhaps the result of applying controls for experience and job status. An alternative hypothesis is that many

of the first jobs are held after three-year apprenticeships and often in the same firms. Therefore they are less likely to show the characteristics of a job search period.

Industrial sector

Social class and industrial sector are investigated simultaneously in Table 3. It reports estimates of any job change for each of Stinchcombe's (1979) industrial sectors.

(Table 3 about here)

According to these estimates, class distinctions are important for job mobility in the capitalist sector (food,wood, leather, textiles,printing) and in the small competitive sector (sales, retail,hotels, restaurants, personal and domestic services). Social class is without statistically significant effects in the professional, bureaucratic and large-scale engineering sectors. As argued in the structural dominance thesis, the more knowledge-based and bureaucratic sectors most likely superimpose their own mobility regimes within internal labor markets.

Job, Sector and Class

How do changes in job or sector interrelate with changes in social class? Table 4 presents conditional probabilities of changing social class, by origin and destination class, given a change in job or industrial sector.

(Table 4 about here)

Notice first the strong bonding power of social class. Most job changes, even when associated with sectoral changes, do not involve changes in social class. This is especially the case for the professional classes (I and II) and the semi- and unskilled manual class (VIIa). Only in the self-employed and manual agricultural classes does a job change typically involve a class change, and both of these types of events are almost true by

class definition. We should add that the overall finding of
stability in class structure is not the well-known
"mover-and-stayer" finding of mobility research. In fact, every
individual in the data used in Table 4 has moved in the sense
that he or she has changed jobs.

For those who do change class at the time of a job change, moves
of a short "distance" are the norm. Noteworthy exceptions to this
rule are the high proportions of moves from the self-employed
class(IV), and from the routine non-manual class (III), into the
semi- and unskilled manual class (VIIa). Overall, however, the
findings are consistent with the closure thesis.

Class Mobility

The analyses above should be familiar to students of job mobility
(see Sorensen and Tuma, 1981; Tuma, 1985; Carroll and Mayer,
1986). We report them to show the constraining power of classes
on jobs but also for contrast with our next set of analyses,
which investigate movement into and out of specific classes. When
this topic has been studied, class mobility is often used as a
proxy for job shifts not available in the data. The research
problem, however, is considerably different. Human capital
theory, for instance, makes a lot of sense with regard to jobs
whereas analogous reasoning with respect to classes seems far
from adequate. Underlying our research here is thus the question:
How similar are patterns of job-shifts and class-shifts?

Table 5 presents partial likelihood estimates of the rate of
leaving any social class as a function of the various independent
variables. The left-most equation reports the base model without
any class independent variables . The next equation to the right
reports a model including the origin classes as dummy variables
(again, the farm workers are the omitted contrast). The remaining
equations report the basic model estimated separately for each
origin class.

(Table 5 about here)

We note several important features of these estimates. First, the strong cohort effects disappear when moving from rates of job shifts to class shifts. Substantively, this finding is consistent with the rationalization thesis: class barriers have not become easier to transcend although mobility between jobs within classes seems to have increased. Methodologically, it remains an open question whether the higher job mobility of the younger cohorts is a true substantive finding or an artifact of the retrospective method. However, the assumption of strong retrospective biases should affect job and class mobility likewise. Since this is not the case the finding of job mobility increasing from the oldest to the youngest cohort is corroborated.

Second, the effects of the sex variable change in sign across two types of analysis. Whereas in the job shifts, women moved more frequently, in the class shifts they move less. This finding is not an artifact of the class compositions of women's jobs, as the equations which control for class origin clearly show.

Third, human capital theory, as exemplified by the labor force experience variable, seems to have few implications for class mobility. Whereas this variable has strong negative effects on job shift patterns, its predictive power diminishes to almost nil in the class shift equations once origin class is taken into account.

Fourth, the effects of first class agree with those of first job. Entry class does not appear to be a fluctuating, transitory state. Rather it exerts considerable holding power and the rate of class shift actually rises once first class has been left. This suggests that the life course thesis does not lead to an adequate image of class mobility during working life.

Fifth, size of firm or organization continues to have strong negative effects on the rates of mobility, job-wise or class-wise. This effect persists in the face of controls for origin class, which supports our earlier conclusion that organizational structures shape careers independently of class and industrial sector (Carroll and Mayer, 1986).

Sixth, class differences in mobility are strong and ordered in a
fashion somewhat that suggested by the closure thesis. The higher
service and professional class (I) exerts the greatest
constraining force, manifested in a rate of movement about 90 %
lower than the farm workers (VIIb). The other classes are even
more consistently ordered than in the job-shift rates reported in
Table 3.

We also searched for interaction effects between cohort and
education on the rates of exits from given classes following the
hypothesis which postulates a loosening of the influence of
education on class mobility over historical time. Except for one
instance we found no such historical trend or change. The
exception is a higher positive rate of exits from class II in the
middle cohorts. This suggests superior upward mobility chances
during the career brought about by favorable returns to education
for the particular path leading from the middle to the higher
levels of the class hierarchy.[5]

Entry into initial class

The equations above are conditional on having previously entered
the labor force and thus having an origin class. The strong
effects of the first class episode suggest that entry into the
first class may be a much different process. It is also the
process where evidence of reproduction theory should be most
clearly visible. For these reasons, we estimated models of
initial class entry separately. Here, however, waiting time is
relatively unimportant. It will measure primarily the level of
education. It is also not clear when the "clock" for such a
waiting time should begin. Thus, we report estimates of logistic
response models for initial class entry. They are found in Table
6.

(Table 6 about here)

As might be expected in Germany (Blossfeld, 1985; König/Müller, 1986) the educational and vocational degrees have a strong impact on the initial class entered. Such certified qualifications act as gateways or elevators to class position. Once educational and occupational training are taken into account, the socio-economic resources of the family of origin - measured here by the occupational prestige of father's job when the respondent was 15 years of age - do not influence additionally the class entry of children.The major exception to this pattern is entry into class IV where superior resources of the family of origin ease entry into self-employment. Likewise a low status of the father keeps children in the working class.[6]

The gender variable shows strong signs of class segregation occuring at time of entry into the labor force. Women are channeled into classes II and III in high proportions while they are much less likely than men to move into all the other classes. Coupled with the strong negative effects of sex on class mobility after entry (Table 5), this strongly suggests that most sex segregation is career-long (see also Blossfeld, forthcoming).

Again, the cohort variables usually fail to reach statistical significance. The exception is the equation for farm workers where the long-term trend of mechanization finds as much expression as the transitory increase of farm work after World War II. Both findings support the position that our earlier estimates regarding cohort effects were not methodological artifacts.

Later Class Entry

Given an initial class, which are the conditions under which mobile individuals enter specific classes? We now turn to this question by examining partial likelihood estimates of models with class-specific destination states. That is, we model the destination class as a competing risk model. The estimates are presented in Table 7.

(Table 7 about here)

Entry into the class of higher administrative, professional and managerial positions strongly depends on higher formal schooling and is highly restricted for women. Although the number of such positions has clearly increased over time we detect no cohort effect. One must, therefore, assume that the cohort differences are compositional and hidden in the effect of schooling.

The class II of qualified, semi-professional white collar positions is, in contrast, increasingly accessible to the younger cohorts and is equally open for women and men. General schooling is important and there is a negative effect of occupational status, i.e. moves into class II tend to be status-equilibrating and start from lower jobs.

Class III of lower white-collar work is a destination port for members of the youngest cohort and for women, with little or no vocational training. It is not a condition typically to be entered after the first class episode.

Becoming self-employed (class IV) tends to occur later in the career and is more likely the smaller the size of the employing firm. Thus, self-employment appears to be one way of overcoming the limited opportunities within small firms.

Class IV (technicians, foremen, production supervisors) is a domain of men and a career step for skilled manual workers early in the working life. How does one get into skilled work during the career? This is somewhat of an enigma, since this type of move should usually occur immediately after an apprenticeship. According to these results this move is more prevalent for men, with little training and with destination jobs of lower occuational prestige.

Finally, semi- and unskilled jobs - as a new class - are more likely to be entered by men, by people who come from small firms, with little education and little vocational training.

We also estimated such models for moves into second and subsequent classes including father's occupational prestige as a proxy measure for the socio-economic resources of the family of origin. Such delayed effects of origin not mediated by education shows to be significant only for later moves into class II.[7] Therefore we find that the thesis of class reproduction by means of the educational system is supported in an even stronger manner than suggested above. This mechanism not only operates, as we expected, at the time of entry into the class structure but also during the working life.

Jobs Within Classes

We now take up our last but intriguing research question: How can the number of jobs within a class episode be explained? Recall that our argument, the rationalization thesis, held that job-wise differentiation within classes has increased steadily across the twentieth century while class mobility has not. Our inability to find evidence of cohort effects in the rates of class mobility suggests that part of this thesis may be correct. We now examine the other, more interesting, part directly.

Our research strategy was to estimate ordinary least squares regression models with the number of jobs within the class episode, and its natural logarithm, as dependent variables. Such an approach makes it necessary, of course, to control for duration in the class episode and whether or not the episode was censored. The resulting estimates can be found in Table 8.

(Table 8 about here)

Across the hierarchical models estimated we find very consistent results:

First, in all equations there are strong cohort effects signalling a trend of an increasing number of jobs per class episode over time. Second, the number of jobs is generally higher in the first class spell. This is not an effect of the longer average duration of the first class episode since this is

partialled out. There is therefore a genuine situation of higher
job mobility early in the career. Third, the experience variable
bears out the long established finding of decreasing job mobility
during the career.

More interesting are the findings for the various classes. By
definition, self-employment reduces job mobility. The classes
with the highest partial effect on number of jobs are the two
manual classes of skilled and unskilled workers and the lower
service class (II). They are different in the sense that job
mobility among blue collar workers is most likely horizontal
mobility for small wage gains, whereas in class II career-like
job sequences will be the rule. We do find here a curious
difference between blue collar and white collar positions without
qualification (III,VII). In the unqualified blue collar class job
moves are frequent. They do not appear to be particularly
pronounced in the (mostly female) lower white collar sector. We
do not have a ready-made explanation for this difference in labor
market structure.

CONCLUSION

In this study we introduced two novel aspects into the analysis
of changes during the working life. First, we distinguish both
conceptually and empirically between job shifts and class
mobility. Second, we apply stochastic models for event histories
to class mobility. We believe that the results demonstrate the
great fruitfulness of both steps.

Initially we raised a number of issues of a theoretical and
empirical kind which will now be discussed in the light of our
findings:

1) <u>Does the distinction between job and class mobility matter?</u>
Yes, indeed it does. Although the average number of jobs per
class episode is with 1.54 not particularly striking, most job
<u>changes</u> - even when connected to sectoral shifts - do not involve
changes in social class. This pattern becomes even more

pronounced when we exclude self-employment where job changes almost always imply a move to another (wage earning) class. Also, the gross number of jobs per class episode (Table 1) or the net explanatory power of particular classes (Table 8) show large differences between classes in exposing their members to job shifts. These results confirm our expectations about the higher job mobility of un- and semiskilled and - to a lesser degree - of skilled workers. Unexpected is, however, that the qualified clerical and semi-professional employees experience more job shifts than the routine non-manual employees.

Further, this distinction brings out quite clearly the different labor market experiences of men and women. Women enter into the un- or less-qualified service classes III and II, they experience greater job mobility than men but less class mobility, i.e. they are locked into the more disadvantaged positions.

The distinction between jobs and classes has also proven to be useful in the assessment of inter-cohort changes. In contrast to increases in job mobility from the older to the younger cohorts we find stability in class mobility. This lends credibility to our assumption that the cohort differences in job shifts are not an artifact of the retrospective method. Substantively, we take this to be evidence for an increase in the degree of job differentiation (rationalization hypothesis).

2) <u>Is the class structure a central, macro-social condition for shaping job trajectories?</u>
Particularly in the German case a societal class structure is not only a derivative of the economic order, but is, in addition, on a national level patterned and modified by legal provisions of labor law, social security law, agreements of collective bargaining and the institutions of general education and vocational training. We find abundant evidence that the determining force of class membership does not vanish once we take industrial sector, size of firm and individual resources into account.

However, class effects on either the rate of job change or the
rate of class change do not spread evenly. They are definitely
modified according to industrial sector, whether the moves are
within- or across-firms and according to size of the employing
organization. The less the firm or sector conforms to the market
model, the less is the influence of the class structure on
working lives. Conversely, job trajectories in "systems of open
positions" (Sorensen, 1983) are not governed exclusively by
attributes of the person and the job. The class structure
intervenes (structural dominance thesis).

The impact of the class structure on career mobility may probably
be captured most adequately by an image of inertia. Classes hold
people and predetermine their life space to a considerable
extent. Against the life-course-thesis we do not find that first
class episodes are of shorter duration. The opposite appears to
be the case and this is extremely so if the the entry classes are
the ones of the lower service class(II) and the unskilled workers
(VIIa). 40 % of our respondents stay in their initial class
during the time we observed our c. 30-,40- and 50-year old
respondents retrospectively.

This mobility quota seems to contradict the image of inertia and
self-reproduction. We, therefore, have to add a number of
qualifications. Class VII b of the farm workers is a highly
transitory state where only 17% of our respondents stay their
whole working life and almost 60% move to unskilled jobs.
Likewise the class V of manual supervisors and lower technicians
has many properties of a "career stage" which most persons enter
from skilled manual jobs and almost 50% leave again for
qualified or unqualified white collar positions or
self-employment. Also the majority of men in class III experience
upward mobility to class II or self-employment. Even of the men
ever having been in class II a third make their way up to class
I. Class closure is highest in the most privileged classes I and
II and lowest among the skilled workers (VI). Thus, we have to
stress not only the long mean duration in a given class (about
16-17 years) but also the fact that origin class strongly
patterns the direction and destination of a class change.

FOOTNOTES

1. We abandonned the class scheme by Wright since already in Carroll/Mayer we had to modify it by separating blue and white collar positions and by distinguishing skilled workers from un- and semi-skilled workers. Even in that form it includes a highly heterogeneous class of managers comprising all employed workers exercising control.

2. Since the sample comprises respondents only up to age 53 we have not taken up here the complexities of life course changes in career patterns in the pre-retirement phase. We would expect here considerable amounts of downgrading and downward mobility for older manual workers.

3. For more detailed analyses of long term changes of the occupational structure on career mobility, see Blossfeld (1985, 1986 and forthcoming).

4. This results contrasts with the estimates of models - not reported here - where class conditions are left out as predictors of the rate of job change or of class change. In such models occupational status shows significant effects.

5. The beta coefficient of the interaction between education and the birth cohorts of 1939-41 is .5424 (standard error .273).

6. If one estimates these models without the education and
   training variables, all origin effects except for entry into
   classes III, IV and VI become highly significant. Also the
   changes in educational composition show up in differences
   between cohorts.

7. The beta coefficient for the effect of father's prestige on
   the rate of entry into class II is .0074 (standard error .003).

References

Asendorf-Krings I, Drexel I, Nuber C. (1976): 'Reproduktionsvermögen und die Interessen von Kapital und Arbeit. Ein Beitrag zur theoretischen Bestimmung von Qualifikation, pp. 207-236, Frankfurt am Main: Aspekte Verlag.

Averitt R. (1968): The dual economy, New York: McGraw Hill.

Baron J, Bielby W. (1980): 'Bringing the firms back in: stratification and the economic segmentation of work', American Sociological Review, 45: 737-765.

Baron J, Bielby W. (1984): 'The organization of work in a segmented economy', American Sociological Review, 49: 454-473.

Bell D. (1973): The coming of post-industrial society, New York: Basic.

Bernard P, Renaud J. (1976): 'Contre-mobilité et effets différés', Sociologie et sociétés, 8

Bertaux D. (1974): 'Mobilité sociale biographique: une critique de l'approche transversale', Revue française de sociologie, 15: 329-362.

Blau P, Duncan O. (1967): The American occupational structure, New York: Wiley.

Blossfeld, H-P. (1985): 'Berufseintritt und Berufsverlauf - Eine Kohorten-analyse über die Bedeutung des ersten Berufs in der Erwerbsbiographie', Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, 2: 177-197.

Blossfeld, H-P, Hamerle A, Mayer K U. (1986): Ereignisanalyse. Statistische Theorie und Anwendung in den Wirtschafts- und Sozialwissenschaften, Frankfurt am Main: Campus Verlag.

Blossfeld H-P. (1986a): 'Career opportunities in the Federal Republic of Germany: a dynamic approach to the study of life-course, cohort, and period effects', European Sociological Review, 2 (3): 208-225.

Blossfeld H-P. (1986b):'Zur Repräsentativität der Sfb-3-Lebensverlaufs-studie - Ein Vergleich mit Daten aus der amtlichen Statistik', Allge-meines Statistisches Archiv (forthcoming).

Blossfeld H-P.(forthcoming): 'Labor market entry and the sexual segregation of careers in the Federal Republic of Germany', American Journal of Sociology.

Bourdieu P, Passeron J C. (1971): Die Illusion der Chancengleichheit. Untersuchungen zur Soziologie des Bildungswesens am Beispiel Frankreichs, Stuttgart: Klett Verlag.

Braverman H. (1974): Labor and monopoly capital, New York: Monthly Review Press.

Breiger R. (1981): 'The social class structure of occupational mobility', American Journal of Sociology, 87:578-611.

Brückner E, Hoffmeyer-Zlotnik J, Kirsch H-P, Wiedenbeck M. (1984): Methodenreport 'Lebensverläufe', Mannheim: ZUMA (Technischer Bericht Nr. T84/08).

Carroll G, Mayer K U. (1984): 'Organizational effects in the wage attainment process', The Social Science Journal, 21 (3): 5-22.

Carroll G, Mayer K U. (1986): 'Job-shift patterns in the Federal Republic of Germany: the effects of social class, industrial sector, and organizational size', American Sociological Review, 51 (3): 323-341.

Cox D. (1972): 'Regression models and life tables', Journal of the Royal Statistical Society, B 34: 187-220.

Cox D. (1975): 'Partial likelihood', Biometrika, 62: 269-276.

Crompton R. (1980): 'Class mobility in modern Britain', Sociology, 14: 117-119.

Crompton R, Jones G. (1984): White-collar proletariat: deskilling and gender in clerical work, London: Macmillan.

Dahrendorf R. (1969): Class and class conflict in industrial society, London: Routledge and Kegan Paul.

Efron B. (1977): 'The efficiency of Cox's likelihood function for censored data', Journal of the American Statistical Association, 72: 557-565.

Erikson R, Goldthorpe J. (1985): 'Are American rates of social mobility
    exceptionally high? New evidence on an old issue', European Sociological
    Review, 1 (1): 1-22.

Featherman D, Selbee K. (1986): 'Class formation and class mobility: a
    new approach with counts from life history data'. Paper presented
    at the 1986 Annual Meeting of the American Sociological Association,
    New York.

Giddens A. (1973): The class structure of the advanced societies, London:
    Hutchinson.

Girod R. (1971): Mobilité sociale: faits établis et problèmes ouverts,
    Geneva: Droz.

Goldthorpe J. (1980): Social mobility and class structure in modern Britain,
    Oxford: Clarendon Press.

Goldthorpe J, Payne C. (1986): 'Trends in intergenerational class mobility
    in England and Wales, 1971-1983', Sociology, 20: 1-20.

Haller M, Hodge R. (1981): 'Class and status as dimension of career mobility.
    Some insights from the Austrian case', Zeitschrift für Soziologie, 10:
    133-150.

Hernes G. (1976): 'Structural change in social processes', American Journal
    of Sociology, 82: 513-547.

Kern H, Schumann M. (1970): Industriearbeit und Arbeiterbewußtsein, Frankfurt
    am Main: Europäische Verlagsanstalt.

König W, Müller W. (1986): 'Educational system and labour markets as determi-
    nants of worklife mobility in France and West Germany: a comparison of
    men's career mobility, 1965-1970', European Sociological Review, 2 (2):
    73-96.

Kosta J, Krings I, Lutz B. (1970): 'Probleme der Klassifikation von Erwerbs-
    tätigen und Tätigkeiten', München: Institut für sozialwissenschaftliche
    Forschung.

Lappe L. (1985): 'Berufsverlaufsmuster und Reproduktionsinteressen junger Facharbeiter', in E H Hoff, L Lappe, W Lempert (eds), Arbeitsbiographie und Persönlichkeitsentwicklung, pp. 179-199, Bern: Huber.

March J C, March J G. (1981): 'Performance sampling and Weibull distributions' Administrative Science Quarterly, 26: 90-92.

Marx K. (1934): The class struggles in France (1848-1850), New York: International Publishers.

Marx K. (1914-1919): Das Kapital. Kritik der politischen Ökonomie (vol. 1), Hamburg: Meissner.

Marx K, Engels F. (1970): Werke (vol. 23), Berlin: Dietz.

Mayer K U, Müller W. (1972): 'Die Analyse von Mobilitätstrends', Kölner Zeitschrift für Soziologie und Sozialpsychologie, 24: 132-139.

Mayer K U. (1977): 'Ungleiche Chancen und Klassenbildung', Soziale Welt 28: 466-493.

Mayer K U. (1979a): 'Berufliche Tätigkeiten, berufliche Stellung und beruflicher Status - empirische Vergleiche zum Klassifikationsproblem', in F U Pappi (ed), Sozialstrukturanalyse mit Umfragedaten, pp. 79-123, Königstein: Athenäum.

Mayer K U. (1979b): 'Strukturwandel im Beschäftigungssystem und berufliche Mobilität zwischen Generationen', Zeitschrift für Bevölkerungswissenschaft, 3: 267-298.

Mayer K U. (1986): 'Structural constraints on the life course', Human Development, 29: 163-170.

Mayer K U, Wagner M. (1986): 'Der Auszug von Kindern aus dem elterlichen Haushalt - ein Erklärungsmodell für die Geburtsjahrgänge 1929-31, 1939-41 und 1949-51', in K F Zimmermann (ed), Demographische Probleme der Haushaltsökonomie, pp. 1-42, Bochum: Studienverlag Dr. N. Brockmeyer.

Müller W. (1982): 'Wege und Grenzen der Tertiarisierung: Wandel der Berufsstruktur in der Bundesrepublik Deutschland 1950-1980', in J Matthes (ed), Krise der Arbeitsgesellschaft? Verhandlungen des 21. Deutschen Soziologentages in Bamberg, pp. 142-160, Frankfurt am Main: Campus Verlag.

Oppenheimer V K. (1974): 'The life-cycle squeeze: The interaction of men's occupational and family life cycles', Demography, 11: 227-245.

Papastefanou G. (1980):'Zur Güte von retrospektiven Daten - Eine Anwendung gedächtnispsychologischer Theorie und Ergebnisse einer Nachbefragung', Johann Wolfgang Goethe-Universität, Frankfurt am Main, Universität Mannheim, Sonderforschungsbereich 3, Arbeitspapier Nr. 29.

Parkin F. (1971): Class inequality and political order, London: McGibben and Kee.

Parkin F. (1974): 'Strategies of social closure in class formation', in F. Parkin (ed), The social analysis of class structure, pp.1-18, London: Tavistock.

Peterson T, Spilerman S. (1986): 'Departures from an internal labor market'. Paper presented at the International Conference on Applications of Event History Analysis in Life Course Research, Berlin: Max-Planck-Institut für Bildungsforschung.

Roemer J. (1982): A general theory of exploitation and class, Cambridge, Mass. Harvard University Press.

Schelsky H. (1965): 'Die Bedeutung des Schichtungsbegriffes für die Analyse der gegenwärtigen deutschen Gesellschaft', in H. Schelsky, Auf der Suche nach Wirklichkeit, Düsseldorf: Diederichs Verlag.

Schelsky H. (1968): 'Die Bedeutung des Klassenbegriffs für die Analyse unserer Gesellschaft', in B. Seidel, S. Jenkner (eds), Klassenbildung und Sozialschichtung, Darmstadt: Wissenschaftliche Buchgesellschaft.

Schumpeter J. (1953): 'Die sozialen Klassen im ethnisch homogenen Milieu', in J. Schumpeter, Aufsätze zur Soziologie, Tübingen: Mohr/Siebeck.

Sengenberger W. (1975): Arbeitsmarktstruktur - Ansätze zu einem Modell des segmentierten Arbeitsmarkts, Frankfurt am Main, München: Aspekte Verlag.

Snipp M. (1985): 'Occupational mobility and social class: insights from men's career mobility', American Sociological Review, 50: 475-493.

Sørensen A B. (1974): 'A model for occupational careers', American Journal of Sociology, 80 (1): 44-57.

Sørensen A B. (1983): 'Processes of allocation to open and closed positions in social structure', Zeitschrift für Soziologie, 12 (3): 203-224.

Sørensen A B, Tuma N B. (1981): 'Labor market structures and job mobility', Research in Social Stratification and Mobility, 1: 67-94.

Stinchcombe A. (1979): 'Social mobility in industrial labor markets', Acta Sociologica, 22: 217-245.

Tölke A. (1980): 'Zuverlässigkeit retrospektiver Verlaufsdaten - Qualitative Ergebnisse einer Nachbefragung', Frankfurt am Main, Mannheim: Johann Wolfgang Goethe-Universität, Universität Mannheim, Sonderforschungsbereich 3, Arbeitspapier Nr. 30.

Tuma N B, Hannan M. (1984): Social dynamics: models and methods, New York: Academic Press.

Tuma N B. (1985): 'Effects of labor market structure on job-shift patterns', in J. Heckman , B. Singer (eds), Longitudinal analysis of labor market data, Cambridge, Mass.: Cambridge University Press.

Weber M. (1964): Wirtschaft und Gesellschaft, Köln and Berlin: Kiepenheuer & Witsch.

Weber M. (1978): Economy and society, Berkeley: University of California Press

Wegener B. (1985): 'Gibt es Sozialprestige?', Zeitschrift für Soziologie, 14: 209-235.

Wright E O. (1979): Class structure and income determination, New York: Academic Press.

APPENDIX A

INDUSTRY SURVEY RESPONSE CODES

(TRANSLATED FROM GERMAN)

---

01 - AGRICULTURE, FORESTRY, ANIMAL HUSBANDRY, FISHING

02 - ENERGY, WATER, MINING
PUBLIC UTILITIES

03 - CHEMICALS (INCL. OIL), STONES, EARTH, GLASS, RUBBER AND ASBESTOS

04 - IRON, STEEL, NON-FERREOUS METALS

05 - MACHINERY, CARS, STEEL CONSTRUCTION

06 - OFFICE MACHINES (INCL. COMPUTERS)

07 - ELECTROTECHNICAL, METALS, MECHANICS, OPTICS, MUSICAL AND
SPORT INSTRUMENTS, TOYS AND JEWELLERY

08 - WOOD, PAPER, PRINTING, LEATHER, TEXTILES (INCL. CLOTHES)

09 - FOOD

10 - CONSTRUCTION

11 - SALES, RETAIL

12 - COMMUNICATIONS, TRANSPORTATION

13 - BANKING AND INSURANCES

14 - HOTELS, RESTAURANTS

15 - PERSONAL HYGIENE (E.G. HAIRDRESSER)

16 - SCIENCE, EDUCATION, ARTS, PRESS (E.G. SCHOOLS AND THEATRE)

17 - HEALTH

18 - LAW, TAX ACCOUNTANTS, ENGINEERING, REAL ESTATE, PHOTOGRAPHY
AND OTHER SERVICES

19 - OTHER PRIVATE SERVICES

20 - CHURCHES, ASSOCIATIONS, PRIVATE HOUSEHOLDS

21 - STATE OFFICES, DEFENSE, SOCIAL INSURANCE

APPENDIX B

MAPPING OF SURVEY RESPONSE CODES INTO INDUSTRY SCHEMA

01 TRADITIONAL PRIMARY

02 LARGE-SCALE ENGINEERING

03 LARGE-SCALE ENGINEERING

04 LARGE-SCALE ENGINEERING

05 LARGE-SCALE ENGINEERING

06 LARGE-SCALE ENGINEERING

07 COMPETITIVE

08 CLASSICAL CAPITALIST

09 CLASSICAL CAPITALIST

10 COMPETITIVE

11 SMALL COMPETITIVE

12 LARGE-SCALE ENGINEERING

13 BUREAUCRATIC

14 SMALL COMPETITIVE

15 SMALL COMPETITIVE

16 PROFESSIONAL

17 PROFESSIONAL

18 PROFESSIONAL

19 SMALL COMPETITIVE

20 PROFESSIONAL

21 BUREAUCRATIC

Table 1: Distribution of Social Class in West Germany Across the Life Course

Social Class	All Jobs	All Jobs by Sector			First Jobs	First Jobs by Cohort			All Class Episodes	Mean Jobs per Class Episode
		Large-Scale Engeneering	Classical Capitalist	Professional		Cohort 1929-31	Cohort 1939-41	Cohort 1949-51		
I. Higher professional, administrative and managerial	273	68	16	116	53	11	15	27	180	1.52
II. Lower professional, administrative and managerial	900	131	55	324	229	46	67	116	528	1.70
III. Routine non-manual	943	143	103	201	337	100	127	150	627	1.50
IV. Small employers, proprietors and self-employed	345	16	30	9	113	60	37	16	285	1.21
V. Lower technical and manual supervisory	422	152	74	29	65	9	25	31	286	1.48
VI. Skilled manual	1514	574	299	35	591	162	227	202	795	1.90
VIIa. Semi- and unskilled manual, non-agricultural	1605	460	409	164	378	169	130	79	1185	1.35
VIIb. Semi- and unskilled manual, agricultural	197	0	4	0	115	86	24	5	146	1.35
Total	6199	1544	990	878	1921	643	652	626	4032	1.54

Table 2: Partial Likelihood Estimates of Effects of Social Class on Rates of Different Types of Job Change
(Standard Errors of Estimates Shown in Parentheses)

	Any Move	Upward Moves	Lateral Moves	Downward Moves	Within Firm	Across Firms	Upward Within Firm	Lateral Within Firm	Upward Across Firms	Lateral Across Firms
Cohort 1939-41	.151* (.038)	.064 (.070)	.077 (.082)	.377* (.104)	.055 (.093)	.157* (.045)	-.026 (.193)	-.005 (.188)	.090 (.075)	.087 (.093)
Cohort 1949-51	.236* (.044)	-.076 (.082)	.208* (.091)	.540* (.117)	.204 (.107)	.147* (.052)	.096 (.223)	.331 (.200)	-.083 (.088)	.160 (.105)
Sex	.224* (.038)	.135 (.074)	.009 (.087)	.728* (.100)	.470* (.100)	.045 (.046)	-.314 (.217)	.578* (.211)	.194* (.079)	.166 (.098)
Experience	-.005* (.000)	.007* (.001)	-.005* (.001)	-.004* (.001)	-.003* (.001)	-.006* (.000)	-.003* (.002)	-.003* (.001)	-.008* (.001)	-.005* (.001)
First Job	-.206* (.042)	.020 (.077)	-.285* (.090)	-.768* (.117)	.253* (.104)	-.328* (.050)	.252 (.214)	.163 (.200)	-.007 (.084)	-.443* (.103)
Job Status	-.0002 (.0015)	-.003 (.003)	.003 (.003)	.005 (.004)	.001 (.003)	.001 (.002)	-.006 (.007)	.008 (.006)	-.002 (.003)	.002 (.004)
Education	.213* (.038)	.279* (.070)	.171* (.078)	.112 (.098)	.275* (.089)	.200* (.046)	.534* (.181)	.513* (.164)	.243* (.076)	.041 (.093)
Training	-.047 (.027)	-.090 (.053)	.008 (.056)	-.195* (.073)	.094 (.061)	-.102* (.034)	.037 (.130)	.194 (.110)	-.118* (.059)	-.021 (.067)
Lnsize Org.	-.060* (.008)	-.117* (.015)	-.032* (.016)	-.016 (.019)	.057* (.018)	-.098* (.010)	.001 (.037)	.145* (.034)	-.140* (.016)	-.084* (.019)
Class I	-.955* (.160)	-1.25* (.310)	-.629 (.389)	-1.13* (.542)	-1.19* (.378)	-1.01* (.197)	-.659 (.874)	-2.08* (.793)	-1.40* (.348)	-.339 (.451)
Class II	-.444* (.105)	-.751* (.192)	.112 (.306)	-.187 (.378)	-.250 (.275)	-.455* (.124)	-.113 (.691)	-.869 (.653)	-.807* (.205)	.300 (.347)
Class III	-.200* (.091)	-.331* (.162)	.384 (.292)	.073 (.360)	-.135 (.261)	-.216* (.107)	.318 (.649)	.155 (.625)	-.364* (.169)	.393 (.331)
Class IV	-1.03* (.120)	-1.94* (.289)	-1.44* (.434)	.679 (.423)	-.589 (.305)	-1.09* (.142)	-1.99 (1.17)	-.917 (.833)	-1.90* (.299)	-1.65* (.515)
Class V	-.399* (.106)	-.654* (.192)	.250 (.301)	-.152 (.387)	-.459 (.276)	-.370* (.124)	.000 (.671)	-.222 (.639)	-.684* (.206)	.346 (.343)
Class VI	-.050 (.085)	-.307* (.144)	.596* (.278)	.403 (.349)	-.153 (.238)	-.015 (.098)	.324 (.607)	.004 (.602)	-.331* (.151)	.716* (.314)
Class VIIa	-.006 (.084)	-.181 (.144)	.511 (.281)	.460 (.347)	.057 (.241)	.010 (.097)	.503 (.611)	.339 (.611)	-.220 (.150)	.543 (.317)
Chi Square	794.9	568.1	184.0	236.2	147.8	630.6	47.4	94.7	564.1	173.4
Degrees of Freedom	16	16	16	16	16	16	16	16	16	16

*p ≤ .05

Table 3: Partial Likelihood Estimates of Social Class Models of the Rate of Any Job Change by Industry Sector
(Standard Errors of Estimates Shown in Parentheses)

	Primary	Large-Scale Engineering	Bureaucratic	Professional	Small Competitive	Capitalist	Competitive
Cohort 1939-41	.024 (.150)	.206* (.080)	.237 (.163)	-.019 (.111)	.159 (.102)	.074 (.086)	.167 (.101)
Cohort 1949-51	.442 (.271)	.293* (.092)	.211 (.159)	.081 (.118)	.363* (.110)	.389* (.102)	.117 (.122)
Sex	-.025 (.130)	.373* (.109)	.087 (.133)	.357* (.125)	.394* (.097)	.014 (.096)	.348* (.157)
Experience	-.011* (.002)	-.005* (.001)	-.006* (.001)	-.005* (.001)	-.004* (.001)	-.004* (.001)	-.004* (.001)
First Job	-.803* (.190)	-.054 (.088)	-.210 (.146)	-.184 (.112)	.115 (.105)	-.336* (.097)	-.179 (.116)
Job Status	-.010 (.008)	.006 (.004)	-.005 (.006)	-.000 (.003)	.002 (.004)	.010* (.004)	.001 (.005)
Education	.734* (.352)	.251* (.081)	.182 (.104)	.056 (.096)	.373* (.094)	.126 (.110)	.294* (.104)
Training	-.370 (.196)	.056 (.069)	-.139 (.100)	-.078 (.057)	-.016 (.078)	-.155 (.097)	.142 (.105)
Lnsize Org.	.066 (.068)	-.047* (.015)	-.071* (.029)	-.082* (.026)	.002 (.027)	-.071* (.022)	-.040 (.023)
Class I	-1.51 (1.16)	-1.01 (.764)	-1.38* (.662)	-.665 (.492)	-3.05* (.715)	-1.61* (.590)	-1.63* (.817)
Class II	1.84* (.847)	-.281 (.725)	-.402 (.474)	-.301 (.460)	-1.54* (.477)	-1.52* (.464)	-1.21 (.744)
Class III	1.83* (.735)	.088 (.721)	-.349 (.482)	-.408 (.466)	-1.15* (.457)	-1.00* (.434)	-.948 (.762)
Class IV	-.726* (.208)	-1.25 (.918	.784 (.740)	-1.41 (.741)	-1.90* (.492)	-2.37* (.521)	-1.92* (.829)
Class V	-.209 (.530)	-.006 (.718)	-.810 (.511)	-.196 (.517)	-1.11* (.486)	-1.09* (.443)	-1.21 (.750)
Class VI	.469 (.429)	.399 (.711)	-.396 (.540)	-.104 (.503)	-1.17* (.461)	-.903* (.418)	-.407 (.739)
Class VIIa	.007 (.235)	.487 (.714)	-.199 (.527)	-.255 (.479)	-.914* (.462)	-.771 (.421)	-.406 (.739)
Chi Square	125.6	201.6	85.0	114.7	148.4	114.8	102.2
Degrees of Freedom	16	16	16	16	16	16	16
Number of Jobs	308	1290	490	717	824	890	729

*p ≤ .05

Table 4: Conditional Probabilities of a Change in Social Class Given a Change in Job or Sector

Origin Class	Destination Class								Total from Origin
	I J/S	II J/S	III J/S	IV J/S	V J/S	VI J/S	VIIa J/S	VIIb J/S	
I	82.1/77.3	4.5/4.5	2.7/4.5	8.1/9.0	0/0	1.8/4.5	0.9/0	0/0	100/100
II	11.2/6.9	73.2/62.1	9.3/17.8	2.2/4.0	1.0/1.1	0.4/1.1	2.8/6.9	0/0	100/100
III	1.3/1.4	18.3/17.9	57.8/50.3	4.7/5.2	1.6/2.1	2.9/3.4	13.2/19.3	0.2/0.3	100/100
IV	2.8/2.2	6.1/14.3	6.7/12.1	36.1/5.5	3.9/2.2	7.2/13.2	28.3/50.5	5.6/0	100/100
V	5.6/6.8	13.2/17.8	6.0/11.0	9.3/6.8	58.1/46.6	5.6/6.8	2.1/4.1	0/0	100/100
VI	1.5/0.9	3.6/5.5	2.6/6.4	3.1/3.6	13.1/8.8	62.1/41.9	13.9/32.5	0.1/0.3	100/100
VIIa	0.6/0.6	2.8/3.1	7.1/9.4	2.9/4.1	3.1/1.1	10.5/10.9	71.3/68.3	1.7/2.4	100/100
VIIb	0/0	1.2/1.9	2.9/4.9	8.8/1.0	1.8/1.9	7.0/11.7	48.5/76.7	28.8/1.9	100/100

Note: J/S refers to a conditional job change and sector change respectively.

Table 5: Partial Likelihood Estimates of Models of the Rate of Class Change
(Standard Errors Shown in Parentheses)

Independent Variables	All Origins	All Origins	Origin Class I	Origin Class II	Origin Class III	Origin Class IV	Origin Class V	Origin Class VI	Origin Class VIIa	Origin Class VIIb
Cohort 1939-41	.069 (.067)	.081 (.067)	-.012 (.619)	-.315 (.215)	-.105 (.196)	.498 (.265)	-.210 (.266)	.058 (.116)	.160 (.152)	-.213 (.247)
Cohort 1949-51	.063 (.079)	.074 (.081)	.146 (.737)	-.684* (.249)	.030 (.208)	.777* (.354)	-.458 (.331)	.036 (.137)	.165 (.188)	.514 (.632)
Sex	-.288* (.065)	-.205* (.071)	-1.58 (1.20)	.108 (.190)	-.369* (.181)	-.566* (.272)	.594 (.361)	.027 (.155)	-.065 (.154)	-.436* (.229)
Education	.296* (.066)	.393* (.069)	.644 (.485)	.316* (.144)	.232 (.147)	.279 (.260)	.681* (.206)	.636* (.151)	1.01* (.180)	.759 (.732)
Training	-.079 (.048)	-.038 (.051)	-.746* (.311)	-.357* (.104)	.109 (.147)	-.319 (.210)	.537* (.193)	.060 (.162)	.423* (.151)	-.069 (.370)
Experience	-.005* (.001)	-.005* (.001)	-.006 (.005)	-.006* (.002)	-.005* (.002)	-.009* (.003)	-.004* (.002)	-.003 (.002)	-.003 (.002)	.003 (.004)
First Class	-.227* (.083)	-.414* (.086)	-.460 (.861)	-.645* (.251)	-.347 (.220)	-.420 (.417)	.125 (.316)	-.212 (.199)	-.328 (.174)	-.836* (.340)
Job Status	-.013* (.002)	-.002 (.003)	-.018 (.012)	.021* (.007)	-.024* (.007)	.022* (.008)	.019 (.010)	.003 (.009)	.003 (.007)	-.021 (.013)
Log.Org.Size	-.086* (.014)	-.065* (.015)	-.311* (.143)	-.052 (.044)	-.065 (.044)	-.071 (.153)	-.028 (.051)	-.069* (.026)	-.032 (.029)	.029 (.156)
Origin Class I		-2.68* (.333)								
Origin Class II		-1.53* (.173)								
Origin Class III		-.854* (.137)								
Origin Class IV		-1.19* (.164)								
Origin Class V		-1.16* (.161)								
Origin Class VI		-.695* (.118)								
Origin Class VIIa		-1.02* (.122)								
Chi Square	168.8	324.4	17.5	34.3	35.7	35.4	53.0	34.9	66.7	28.8
Degrees of Freedom	9	16	9	9	9	9	9	9	9	9

*p ≤ .05

Table 6: Logistic Response Models of Initial Class Entry

Independent Variables	Entry							
	Class I	Class II	Class III	Class IV	Class V	Class VI	Class VIIa	Class VIIb
Constant	-3.75*	-1.36*	-2.11*	-7.80	-5.29	-3.93	-5.47	-7.05
	(.509)	(2.87)	(.378)	(n.e.)	(3,51)	(n.e.)	(3.51)	(3.79)
Cohort 1939-41	.049	.138	.064	-.019	.319	.102	.089	-.598*
	(.263)	(.123)	(.088)	(.119)	(.219)	(.092)	(.086)	(.137)
Cohort 1949-51	.086	.180	.148	-.299	.488*	.013	.078	-1.15*
	(.239)	(.116)	(.088)	(.162)	(.209)	(.094)	(.100)	(.263)
Sex	-.385*	.445*	1.16*	-.036	-.342*	-1.12*	-.104	-.412*
	(.196)	(.093)	(.088)	(.112)	(.159)	(.081)	(.080)	(.118)
Father's job prestige	.007	.001	.000	.215*	.010	-.012*	-.005	-.011
	(.005)	(.003)	(.003)	(.006)	(.006)	(.004)	(.004)	(.008)
General Education:								
Middle school degree	-.366	.108	.238	-.415	-.321	-.353	-.170	1.76
	(.551)	(.144)	(.142)	(.491)	(.288)	(.246)	(.225)	(3.76)
High school degree	1.49*	1.57*	.189	-1.37	.426	-1.15*	-.652*	-4.40
	(.523)	(.199)	(.218)	(.740)	(.396)	(.419)	(.316)	(7.49)
Occupational Training:								
Apprenticeship	-1.27*	.196	.912*	.793	1.91	3.02*	1.79	-.114
	(.484)	(.245)	(.341)	(.712)	(3.49)	(.329)	(3.50)	(.158)
Technical training	1.90*	1.62*	1.57*	-3.91*	3.82	.967	-2.37	-4.65
	(.994)	(.690)	(.766)	(1.97)	(3.53)	(.904)	(n.e.)	(n.e.)
Technical school degree	.200	-.377	.369	-2.46	.699	-5.10*	-1.46	1.20
	(.603)	(.433)	(.713)	(n.e.)	(3.58)	(1.69)	(n.e.)	(n.e.)
University degree	1.14*	-.137	-3.00*	2.44	-6.23	1.04	-2.77	1.49
	(.482)	(.303)	(.864)	(1.35)	(13.9)	(1.12)	(6.99)	(n.e.)
Log likelihood	-113.8	-441.3	-674.6	-321.7	-213.5	-607.5	-568.1	-283.9
Chi Square	178.5	575.2	631.9	325.1	226.6	490.8	482.2	185.2
Degrees of Freedom	679	679	679	681	679	680	681	682
No. of Events	46	190	314	105	55	484	295	99

*p ≤ .05

Table 7: Partial Likelihood Estimates of Models of the Rate of Class Entry
(Standard Errors Shown in Parentheses)

Independent Variables	Destination Class						
	I	II	III	IV	V	VI	VIIa
Cohort 1939-41	.025	.456*	.230	-.131	.265	.138	-.025
	(.245)	(.185)	(.200)	(.198)	(.171)	(.209)	(.127)
Cohort 1949-51	-.229	.575*	.443*	.093	.224	-.087	.050
	(.296)	(.204)	(.226)	(.239)	(.211)	(.271)	(.155)
Sex	-1.42*	-.016	1.33*	-.099	-1.61*	-1.04*	-.447*
	(.319)	(.158)	(.194)	(.190)	(.257)	(.244)	(.125)
Education	1.36*	.894*	.249	.151	-.355	-.562	-1.01*
	(.184)	(.133)	(.180)	(.204)	(.220)	(.356)	(.258)
Training	.026	-.182	-.349*	.099	.450*	-.518*	-.268*
	(.118)	(.100)	(.138)	(.140)	(.144)	(.192)	(.119)
Experience	-.005	-.000	-.003*	-.006*	-.003	-.009*	-.005*
	(.003)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)
First Class	.052	.263	-.635*	-.879*	.538*	-1.34*	-.001
	(.321)	(.220)	(.218)	(.229)	(.250)	(.238)	(.171)
Job Status	-.006	-.011*	-.003	-.011	-.025*	-.026*	-.018*
	(.005)	(.005)	(.006)	(.006)	(.007)	(.008)	(.005)
Log. Org. Size	-.005	.020	-.025	-.233*	-.033	.061	-.281*
	(.045)	(.033)	(.039)	(.045)	(.034)	(.040)	(.032)
Chi Square	183.6	72.3	100.3	47.3	117.9	91.4	198.9
Degrees of Freedom	9	9	9	9	9	9	9

*p ≤ .05   Note: Class VIIb had too few observed entries to allow for estimation.

Table 8: Models of Within-Class Differentiation by Job Shifts
(Standard Error Shown in Parentheses)

Independent Variables	Dependent Variable							
	Number of Jobs	Log. Number of Jobs	Number of Jobs	Log. Number of Jobs	Number of Jobs	Log. Number of Jobs	Number of Jobs	Log. Number of Jobs
Constant	.930	.001	1.06	.056	1.08	.016	.709	-.154
Duration	.476* (.028)	.223* (.012)	.429* (.030)	.201* (.013)	.425* (.031)	.201* (.013)	.422* (.030)	.200* (.013)
Censored	.056 (.055)	.024 (.023)	-.069 (.060)	-.037 (.025)	-.091 (.061)	-.041 (.025)	-.073 (.060)	-.034 (.025)
Cohort 1939-41	.274* (.060)	.129* (.025)	.229* (.060)	.108* (.025)	.238* (.061)	.109* (.025)	.244* (.060)	.109* (.025)
Cohort 1949-51	.368* (.064)	.192* (.027)	.235* (.068)	.129* (.028)	.248* (.068)	.130* (.028)	.238* (.068)	.121* (.028)
First Class			.183* (.065)	.097* (.027)	.230* (.080)	.126* (.033)	.234* (.080)	.118* (.033)
Experience			-.157* (.045)	-.074* (.019)	-.179* (.049)	-.087* (.021)	-.136* (.049)	-.067* (.020)
Sex			.015 (.051)	.014 (.022)	-.020 (.054)	.007 (.022)	.011 (.057)	.026 (.024)
Class Spell No.					.050 (.042)	.027 (.018)	.046 (.042)	.025 (.017)
Education					.028 (.044)	.008 (.018)	.069 (.047)	.029 (.019)
Training					-.068* (.032)	-.012 (.014)	-.037 (.034)	-.101* (.014)
Class I							-.011 (.154)	-.017 (.064)
Class II							.230* (.115)	.126* (.048)
Class III							.139 (.107)	.075 (.044)
Class IV							-.278* (.120)	-.125* (.049)
Class V							.070 (.127)	.030 (.053)
Class VI							.366* (.104)	.199* (.043)
Class VIIa							.573* (.105)	.249* (.043)
R^2	.134	.164	.157	.197	.160	.198	.196	.241

*p ≤ .05

Figure 1: Distribution of Jobs per Class Episode
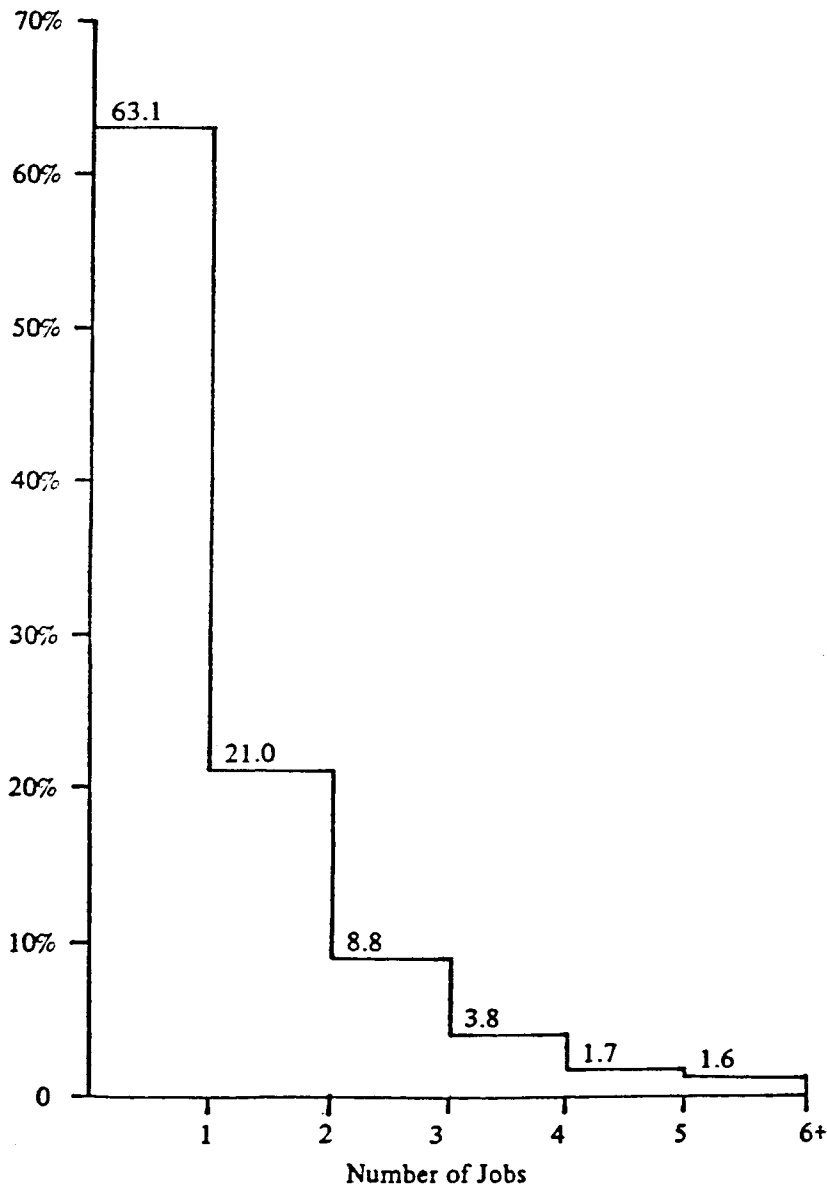
MPI Bilfo
Stat 70/86

Industries, labor markets, firms and occupational careers: On which level does structure matter?

Josef Brüderl

During the last decade research on occupational careers has shifted its focus from individual determinants of socio-economic attainment to the structural determinants such as industry, labor markets or organizations. Baron and Bielby (1980: 737) refer to this work as the 'new structuralism in stratification research'. The new structuralists emphasize the important role of structural factors affecting the individual career process. An impressive body of empirical results supports this proposition (e.g., Beck et al. 1978; Bielby and Baron 1983; Carroll and Mayer 1986; Dickens and Lang 1985; Hodson 1983; Pfeffer 1977; Rosenbaum 1984; Stolzenberg 1978; Tolbert 1982; Wallace and Kalleberg 1981). Yet, there are different lines of research in this tradition. They differ concerning the level of structure to be used (see Baron and Bielby 1980: 743). The oldest line - segmentation theory- chooses the structure of industries respectively labor markets (institutional level). More recently some new structuralists argued for the preeminence

of firm characteristics (organizational level) as determinants of individual careers. Proponents of the third, the job level are yet rarely found (e.g., Spaeth 1984).

There is an ongoing debate about the most fruitful level of structural analysis (see Baron and Bielby 1984; Carroll and Mayer 1986; Hodson 1983; 1984). This paper wants to present some empirical evidence on this question as it examines institutional and organizational determinants of individual career mobility[1]. The results strongly support structural analysis of careers on the firm level. The next section proposes some theoretical arguments for the abandonment of segmentation theories.

Segmentation Theories and Firms

The general argument here is, that segmentation theories are theoretically fruitless, because the effects of labor market segments respectively economic sectors on careers are mediated by firms and the organization of work within these firms (Sørensen (1983) advances a similar argument). This becomes obvious, when one examines the explanations segmentation theorists offer for these effects (for the following see also figure 1).

Adherents of labor market segmentation mostly argue with the existence respectively nonexistence of internal labor markets. Workers in the primary segment, they state, are better paid and got more stable jobs, because they are members of internal labor markets. Doeringer and Piore (1971: 167), who made the concept popular, even postulated an identity of primary segment and internal labor markets. Jobs not organized in closed job-hierarchies are secondary jobs. However, in order to identify internal labor markets one has to look at the firms and the organization of work within them. Thus segmentation is not defined on the market level but on the firm level.

Figure 1: The structure of segmentation theories


```
                        Type of
                      Organization
          ↗                              ↘  Job
                                            Stability
  Sector                                   ↗
          ↘                              ↗
            Segment ←──────────→ Internal
                                Labor-markets
```
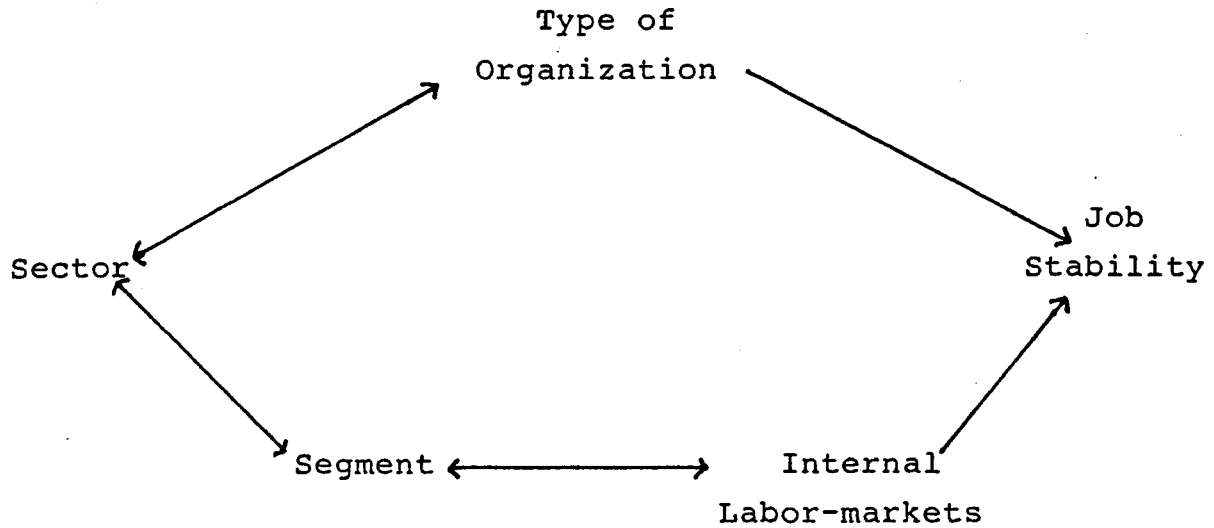
←──→ : correlation

──→ : 'causal'-effect

Similarly, proponents of the 'dual economy' use firm characteristics to explain the effects of sectors on careers (mostly they use them even to define sectors). They maintain that the core contains the bigger and more powerful firms so that the wages are higher. Moreover, there are the big and strong labor unions which improve working conditions and open the 'voice'-option, so that jobs are better and more stable. Clearly, these effects will be captured more directly if one uses the organizational variables themselves. Structural effects on careers are weakened and confounded if one works with sector classifications (see Baron and Bielby 1984). The reason is that sectors are composed of firms with differing features. This heterogeneity is fully accounted for if one looks at the firms themselves. A second explanation of 'dual economists' is based on the assumption, that labor market segments and economic sectors are parallel: In the core there are the

primary jobs, in the periphery the secondary. Thus it is
asserted, that core-jobs are better paid, because they are
mostly primary. But why are primary jobs better paid? The
reasons can be found in the firms as we have seen above.
These considerations show that the effects of segmentation
theories are grounded in firms[2]. Therefore, the more
fruitful strategy for research on careers will be an analy-
sis of the mechanisms that determine careers within organi-
zations. These arguments will be supported by the empirical
analyses on the following pages. It will be shown, that the
effects of segmentation variables on careers vanish if one
controls for firm variables or indicators of internal labor
markets. Hence the latter variables are intervening ones as
implied by the foregoing reflections.

Data

The data used were collected in winter 1980/81 at the
universities of Frankfort and Mannheim under the direction
of Christof Helberger. 2057 people were interviewed. The
sample is representative of all German employees older than
eighteen. The questionnaire included a set of questions
about the interviewee's occupational career (every job-
change with date and some job characteristica were
collected)[3]. The data set contains the duration of every
job-episode. Therefore, it is possible to analyze job
mobility with the methods of survival analysis (see Tuma
and Hannan 1984). I didn't do all-spell analyses, because I
didn't want to confound different career stages and because
of statistical objections (episodes of a single person
aren't independent). The empirical test is carried out with
three different data-sets: the time with the first
employer, the time with the current employer and the
duration for a job-change within the firm. The results for

the first and third data set are presented in the next section. The ones for the second are presented in a separate one, because there are some methodological problems with these data.

Results

The time with the first employer is for every respondent his/her year of the first employer-change minus year of the beginning of his/her occupational career (only year available). Time of job-change within the firm is computed from the difference of job-change year and year of entry into the firm respectively last change. In both cases the risk functions show the same pattern: It starts with relatively low values, reaches its maximum at three years, and declines very fast to lower levels. Figure 2 shows that this pattern is partly due to a German speciality: 'Lehre' (apprenticeship) lasts three years. Therefore, after three years there is a change either within the firm or even one of the employer (see also the effect of 'Lehre' in table 1). But even in the group without 'Lehre' the risk shows its maximum at three years. This non-monotonic pattern would be very hard to parametrize, so I used Cox-regression to analyze both data-sets. Table 1 presents the results. Equations (1) and (3) contain individual variables, dummies for occupational position and the segmentation variable (segment is used, because in these data-sets no information about industry was available).
Let me first say something about the effects of the control-variables. Women leave their first employer faster than men. This is consistent with previous studies (e.g., Carroll and Mayer 1986). The same is true for education. Better educated people are more mobile. They change their employers after a shorter time and are more mobile within

the firm too. Skilled workers and white-collar employees
have a lower risk than unskilled workers in both cases.
Civil-servants show a very low risk of employer change.
Most authors would suggest, that they should have the
highest risk of changing the job within firm, because of
the bureaucratic promotion patterns which prevail for this
occupation. The estimate in equation (3) doesn't confirm
this hypothesis.

Figure 2: The risk of leaving the first employer
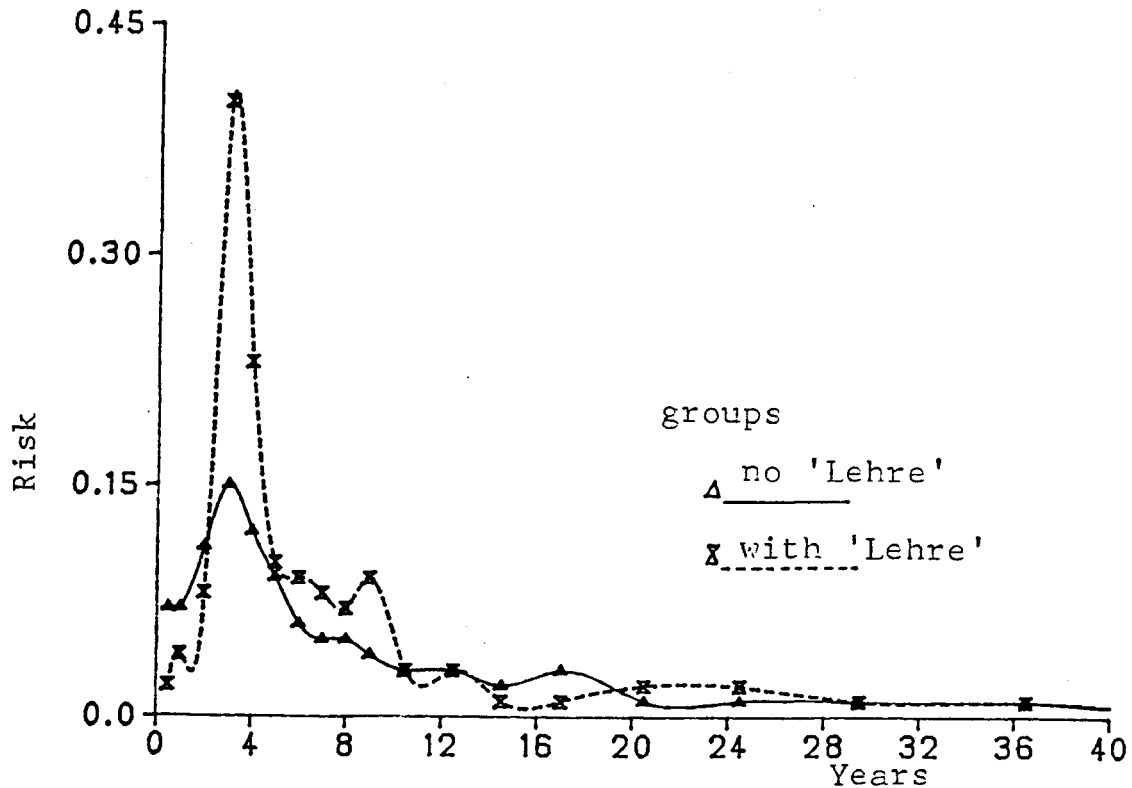
		N	Censored	Median
No	'Lehre'	811	47%	7.7
With	'Lehre'	1073	41%	4.2

Table 1: Cox-regressions with structural variables

	Rate of leaving the first employer		Rate of job-change (within the firm)	
	(1)	(2)	(3)	(4)
Age at beginning of episode	_10	-	-.03* (.005)	-.04* (.005)
Cohort 1949-67[1]	-	-	.34* (.09)	.35* (.10)
Cohort 1968-80[1]	-	-	.36* (.11)	.33 (.11)
Sex[2]	.33* (.09)	.28* (.10)	-	-
Education[3]	.08 (.04)	.10* (.04)	.09* (.02)	.09* (.02)
Part-time job[4]	-1.22* (.51)	-1.18* (.51)	-	-
Apprentice-ship[5] ('Lehre')	.29* (.09)	.26* (.09)	1.30* (.08)	1.31* (.09)
Skilled worker[6]	-.37* (.13)	-.47* (.13)	-.41* (.11)	-.34* (.11)
White collar[6]	-.55* (.15)	-.63* (.15)	-.39* (.12)	-.31* (.12)
Civil-servant[6] ('Beamter')	-2.20* (.31)	-1.87* (.34)	.21 (.14)	-.02 (.16)
Labor market segment[7]	-.27* (.08)	-.06 (.09)	.13 (.07)	.02 (.08)
21-500 employees[8]		-.38* (.09)		.23* (.09)
More than 500 empl.[8]		-.68* (.12)		.25* (.10)
Public-service[8]		-.99* (.19)		.65* (.11)
Chi-square value	114	171	615	630
DF	8	11	9	12
N	1423	1381	5236	5038
% censored[9]	46	46	78	79

* Significant at the 0.05 level.
  Standard errors in parentheses.

1) Year of the beginning of the episode (reference point: 1914-48).
2) Male (0), female (1).
3) Years of schooling + years of vocational training (see Diekmann 1985: 77). In equations (1) and (2) only schooling.
4) Full-time (0), part-time (1).
5) No 'Lehre' (0), 'Lehre' (1) (at the beginning of the episode).
6) Dummies of occupational position; reference point is 'unskilled worker'.
7) Secondary labor market (0), primary labor market (1) (see appendix A for the classification).
8) Dummies of firm size; reference point is 'firms with 1 to 20 employees'.
9) In equations (1) and (2) every respondent is censored, who was at the time of the interview yet with his first employer or who left the job because he had to serve in the army. In equations (3) and (4) every episode that ended with an interview or a change of employer is censored.
10) Covariate not available or with an insignificant effect.

The theory of dual labor markets predicts that workers in the primary segment got more stable jobs than those in the secondary. The estimate in equation (1) fully supports this proposition. On the other side, the theory predicts that promotions within the firm follow in shorter intervals for primary jobs, since these are organized in internal labor markets where promotions are institutionalized. This hypothesis is supported by the coefficient in equation (3) (though it is not significant)[4]. A segmentation theorist would terminate his analysis at this stage and draw the conclusion, that segmentation theory is confirmed by these mobility data. This would, however, be the wrong conclusion as is shown in columns (2) and (4) of table 1. Here I introduce organizational variables. Unfortunately the only information about the firms of the respondents that these data contained ,was the size of the establishment (number of employees; only for workers not in the public service). Previous studies showed, that this variable has a strong effect on employer-stability. It is perhaps the most important of all organizational variables. This is the case, because it seems to be a good indicator of internal labor markets, which yield a strong incentive for the employee to stay with his employer[5]. Nevertheless, size is only a rough indicator of the mechanisms that determine careers within firms. Clearly, if one wants to study these one has to employ much more detailed information about the organizations. But, for our purpose this single firm-variable may suffice (although the test of our hypothesis will be a conservative one). The estimates in equation (2) show, that jobs in larger firms are more stable. The most stable jobs are found in the public-service. And, most important: the effect of the segmentation variable vanishes! The same result do we get from equation (4) (the effects of size are here positive, because internal labor markets further promotions).

Table 2: Cox-regressions with structural variables by labor market segment

	Rate of leaving the first employer		Rate of job-change (within the firm)	
	primary segment	secondary segment	primary segment	secondary segment
Age at beginning of episode	-	-	-.04* (.007)	-.03* (.009)
Cohort 1949-67[1]	-	-	.39* (.14)	.35* (.14)
Cohort 1968-80	-	-	.40* (.15)	.25 (.18)
Sex	.20 (.15)	.33* (.13)	-	-
Education	.11* (.05)	.11 (.09)	.09* (.02)	.09* (.04)
Part-time job	-1.19 (1.01)	-1.13 (.59)	-	-
Apprentice-ship	.21 (.13)	.33* (.14)	1.24* (.11)	1.41* (.14)
Skilled worker	-.46 (.24)	-.55* (.17)	-.17 (.20)	-.47* (.15)
White collar	-.78* (.26)	-.35 (.20)	-.20 (.20)	-.51* (.20)
Civil-servant	-1.95* (.41)	_[2]	-.06 (.23) S	.94* (.35)
21-500 employees	-.25 (.13)	-.49* (.13)	.29* (.13)	.18 (.12)
More than 500 empl.	-.63* (.16)	-.72* (.21)	.36* (.13)	.04 (.18)
Public-service	-.81* (.20)	-1.75 (1.0)	.82* (.14) S	-.11 (.30)
Chi-square value	80	58	386	242
DF	10	9	11	11
N	795	578	2826	2212
% censored	54	33	76	82

* Significant at the 0.05 level.
  Standard errors in parentheses.

1) For a description of the covariates see table 1.
2) In this sample the eight civil-servants were excluded because of estimation problems (see Brüderl 1986: 63).

This clearly confirms our hypothesis: The effects of segmentation are mediated by firms. Therefore, we don't need these variables in our equations on career-mobility. What we need is more subtle information about firms and the organization of work.

A further contention of segmentation theorists is, that the effects of individual covariates (education etc.) are different in different segments. Baron and Bielby (1980) hypothesized, that this is not true for organizational variables. Table 2 presents separate estimations for the segments.

In the case of first employer-change the effects of firm size are relatively similar in the two segments; the order of the coefficients is the same in both segments[6]. In the case of job-change, however, there are differences; size has no effects in the secondary segment, whereas occupational position shows no effects in the primary segment. In both cases the effects for the last dummy are significantly different (this signifies the 'S')[7]. This result is the only one in this paper that gives support for the segmentation theorists. Perhaps mechanisms on the (in this paper neglected) job-level can account for this.

Time with the Current Employer: The Analysis of Backward-Recurrence Times

In this section the results obtained with the third data-set are presented. These data refer to the time one stayed with the current employer. They are backward-recurrence data and at 100% right-censored. Thus, one has to make some considerations about the adequacy of the conventional methods of survival analysis, since these have been developed for duration data that are only partly censored.

The question is, what kind of process we will get, if the
original process is interrupted (for instance by an inter-
view). Sørensen (1977) answers this question for a special
case (following William Feller). He showed, that an
exponentially distributed process, if interrupted, yields
backward-recurrence times which are exponentially distri-
buted too. Thus it is possible to estimate the rate of the
process with the backward-recurrence times and the
constant-rate model is appropriate for the examination of
such data. Allison (1985) extended this analysis and
showed, that the Cox-model too is an appropriate method for
estimating the effects of covariates with backward-
recurrence times. Therefore, I used the Cox-model for the
data on the time with the current employer[8].

However, there are three problems with the causal analysis
of backward-recurrence times.First, backward-recurrence
time data do not contain as much information as ordinary
survival data. This is, because the process that generated
the data is unknown and cannot be infered from the data
themselves. Thus analyses with backward-recurrence times
ought to be considered with some caution. Second, most
covariates are only available for the time of the inter-
view. Thus, if one wants to introduce these variables into
the model, one has to assume, that they didn't change since
the beginning of the process. If this is not true, the
estimates will be biased.

The third problem grounds in the inseparability of the
process of employer-change, which we want to study, and a
second process that results from different job-creation
rates for differing subgroups of the population. An example
clarifies this issue. From the foregoing estimates we know
that civil-servants have the lowest risk of all occupatio-
nal groups to change their employer, that is, they stay
relatively long with their employer. The medians obtained
from the backward-recurrence times show, that this is not
the case for these data:

unskilled workers	11.9 years
skilled workers	14.3 years
white collars	8.3 years
civil-servants	11.2 years

However, it would be wrong to infer that civil-servants have an higher employer-change risk than skilled workers. This is, because the ratio of the job-creation rates of this two groups underwent a great change in the 1970es. Then the public-service expanded rapidly and a great amount of new positions was created. Thus the backward-recurrence times contain relatively more young civil-servants who got their jobs in the 1970es. Therefore the median decreases. The situation is similar for white-collar employees. The estimated effects for 'public-servant' and 'white-collar' would confound both processes. I see no way, how it could be possible to untangle both processes. Thus, one has to consider every covariate as to how important the job-creation effect probably is. For instance, the effects of education or our labor market variable will be seriously distorted. This must be kept in mind when one interprets the following results.

Table 3 is analogous to table 1. It tests whether the effects of the segmentation variable vanish if organizational variables are introduced. For these data information on industry was available. Therefore, a sector-classification was used. Equation (1) shows, that the risk of core-workers to change the employer is lower than the one of periphery workers (albeit not significantly). Public-service workers got the lowest risk (the true effect is probably still stronger, because the job-creation effect makes the risk more positive). Thus, segmentation does well in this equation.

Table 3: Cox-regressions with backward-recurrence times
Time with the current employer

	(1)	(2)
Age at the beginning	.04* (.003)	.03* (.003)
Marital status[1]	.39* (.06)	.37* (.06)
Education[2]	.09* (.01)	.09* (.01)
Earnings[3]	-.0002* (-)	-.0002* (-)
Subordinates[4]	-.14* (.06)	-.12* (.06)
Core[5]	-.08 (.06)	-.006 (.06)
Public-service[5]	-.20* (.08)	-.10 (.08)
Firm size[6]		-.02* (.01)
Changes within firm[7]		-.29* (.03)
Successor from within[8] the firm		-.25* (.06)
Duration of on-the-job[9] training		-.008* (.003)
Chi-square value	350	433
DF	7	11
N	1670	1615
% censored[10]	6	6

* Significant at the 0.05 level.
  Standard errors in parentheses.

1) Married (0), widowed, divorced and single (1) (at the time of the interview (TI)).
2) Years of schooling + years of vocational training (see Diekmann 1985: 77).
3) Monthly, average net income in DM (TI).
4) None (0), at least one (1) (TI).
5) Industry-dummies (TI); reference point is 'peripheral sector' (see appendix B for the classification).
6) Logarithm of number of employees (TI) (categorical variable, midpoint used).
7) Reported number of job-changes at current employer (within the firm).
8) A hypothetical question. Yes (1), No and no answer (0).
9) In months.
10) Censored is every job which began before 1947.

In equation (2) the organizational variables are intro-
duced. For these data more direct indicators of internal
labor markets were available. 'Changes within firm' counts
the number of job-changes the respondent made at the
current employer. The possibility to change jobs is a
necessary condition for internal labor markets. If one is
employed in an internal labor market the successor on the
own job mostly will come from within the firm. The theory
of internal labor markets tells us (see Brüderl 1986:
chapter 3), that firm-specific knowledge is a strong incen-
tive for the firm to install internal labor markets. Thus,
the longer the phase of on-the-job training, the greater
the probability, that the job is integrated in an internal
labor market. These are the three indicators and they show
the expected effects: Clearly, internal labor markets lower
the risk to change the employer. Additionally, firm size
shows the effect we already know. Most importantly, how-
ever, the effects of the sectors almost vanish. This again
supports our hypothesis.

Table 4 contains the regressions separated by sector. Only
three of the nine effects differ significantly (this signi-
fies the 'S'; see footnote 7). That's not much support for
segmentation theories. But again the effects of an organi-
zational variable are differing in the two sectors.
Together with the results in table 2 this shows that
segmentation variables might show some interaction effects.
This, however, is the sole result in this paper by which
segmentation theories got some support.

Table 4: Cox-regressions with backw.-rec. times by industry
Time with the current employer

	core[1] sector		peripheral sector
Age at the beginning	.03* (.004)		.02* (.006)
Marital status[2]	.33* (.07)		.49* (.10)
Education	.07* (.01)	S	.15* (.02)
Earnings	-.0001* (-)	S	-.0003* (.0001)
Subordinates	.17* (.07)		.02 (.12)
Firm size	-.03* (.01)		-.01 (.02)
Changes within firm	-.26* (.04)	S	-.39* (.08)
Successor from within the firm	-.27* (.07)		-.19 (.13)
Duration of on-the-job training	-.005 (.003)		-.02* (.008)
Chi-square value	293		143
DF	9		9
N	1151		464
% censored	6		5

* Significant at the 0.05 level.
  Standard errors in parentheses.

1) Core plus public-service.
2) For a description of the covariates see table 3.

Conclusion

The empirical results presented in this paper demonstrate, that the new structuralists should concentrate their efforts on the organizational level. Controlling for firm variables the effects of the widely used segmentation variables vanished. Segmentation theories were the starting point of the structural approach to career research in the 1970es. The theory was simple and operationalization was easy. Thus, it carried the major burden of the first wave. But now it is time to do a closer look on the processes that build careers. It has been argued in this paper that these processes take place in the firms. Accordingly, research with firm-variables and data on the organization of work are needed.

Notes

1) Job level determinants are not explicitly addressed in this paper. This would be an agenda for future research.
2) The same is true for some modifications of segmentation theories (see Brüderl 1986: 17-21; Sørensen 1983).
3) See Brüderl (1986: 45-47) for further description of the data.
4) It is assumed, that most job-changes are promotions or at least no worsenings (see Rosenbaum 1984: 66).
5) For some other reasons see Brüderl (1986: 37 f).
6) If one ignores the non-significant effect of 'public-service' in the secondary segment. The reason for the great standard error of this estimate is the small number of public-sector workers in the secondary segment.
7) The significance of the difference was tested in a single Cox-regression with interaction-effects for the secondary segment. 'S' signifies that the interaction-effect is significant at the 5% level.
8) Results obtained with the constant-rate model are almost identical. Other authors analyzed backward-recurrence times with OLS-regression (e.g., Bielby and Baron 1983; Freeman 1980). This is possible, because no case is censored (if one ignores the interruption by the interview). Regression results confirm the Cox-estimates in table 3 (see Brüderl 1986: 91f). This demonstrates that the results obtained are very robust against the estimation method used.

Appendix A

Mapping of the ISCO-Classification of occupations into
labor market segments

It was not possible to classify respondents by 'outcomes'
of the stratification process (income, for example, wasn't
available), so their occupations had to be used to assign
them to one of the two segments. Osterman's (1975) classi-
fication couldn't be used, because the survey response
codes differed. So 'the author's judgement' was employed to
map the international ISCO-Classification into the two
segments. The criteria applied were: management occupa-
tions, clerks and professionals were assigned to the prima-
ry segment. Laborers and people in service occupations were
classified with respect to the industry in which they
worked (as far as evident). If they worked in the core they
were assigned to the primary labor market, if they worked
in the periphery they were assigned to the secondary labor
market.

1: secondary labor market
2: primary labor market
9: missing value

segment

001	2	CHEMIKER, PHYSIKER UND VERWANDTE BERUFE
002	2	ARCHITEKTEN, INGENIEURE UND VERWANDTE TECHNIKER
003	2	ARCHITEKTEN, INGENIEURE UND VERWANDTE TECHNIKER
004	2	FLUGZEUG- UND SCHIFFSINGENIEURE
005	2	NATURWISSENSCHAFTLER UND NATURWISSENSCHAFTLICH-
	2	TECHNISCHE HILFSKRAEFTE
006	2	AERZTLICHE, ZAHNAERZTLICHE, TIERAERZTLICHE UND VER-
	2	WANDTE BERUFE
007	2	AERZTLICHE, ZAHNAERZTLICHE, TIERAERZTLICHE UND VER-
	2	WANDTE BERUFE
008	2	STATISTIKER, MATHEMATIKER, SYSTEMANALYTIKER UND VER-
		WANDTE TECHNISCHE SONDERFACHKRAEFTE
009	2	WIRTSCHAFTSWISSENSCHAFTLER
011	2	WIRTSCHAFTSRECHNUNGSSACHVERSTAENDIGE, BUCHPRUEFER
012	2	JURISTEN
013	2	LEHRKRAEFTE
014	2	SEELSORGER, SEELSORGEHELFER
015	2	SCHRIFTSTELLER, JOURNALISTEN UND VERWANDTE PUBLI-
		ZISTISCHE BERUFE
016	2	BILDHAUER, KUNSTMALER, LICHTBILDNER UND VERWANDTE
		GESTALTENDE KUENSTLER
017	2	MUSIKER, DARSTELLER, TAENZER UND AEHNLICHE KUENSTLER
018	2	BERUFSSPORTLER UND VERWANDTE BERUFE
019	2	WISSENSCHAFTLER, TECHNISCHE UND VERWANDTE FACHKRAEF-
		TE, SOWEIT NICHT ANDERWEITIG KLASSIFIZIERT
020	2	ANGEHOERIGE GESETZGEBENDER KOERPERSCHAFTEN UND VER-
		WALTUNGSBEDIENSTETE IN LEITENDER STELLUNG
021	2	FUEHRUNGSKRAEFTE IN DER PRIVATWIRTSCHAFT
030	2	BUEROVORSTEHER
031	2	AUSFUEHRENDE VERWALTUNGSBEDIENSTETE
032	2	STENOGRAPHEN, MASCHINENSCHREIBER, LOCHKARTENLOCHER,
		LOCHSTREIFENLOCHER
033	2	BUCHHALTER, KASSIERER UND VERWANDTE BERUFE
034	2	BEDIENER VON RECHENANLAGEN
035	2	AUFSICHTSKRAEFTE IM TRANSPORT-, FUNK- UND FERNSPRECH-
		WESEN
036	2	SCHAFFNER
037	2	POSTVERTEILER
038	2	TELEPHONISTEN UND TELEGRAPHISTEN
039	2	BUEROKRAEFTE UND VERWANDTE BERUFE, SOWEIT NICHT
		ANDERWEITIG KLASSIFIZIERT
040	2	GESCHAEFTSFUEHRER (GROSSHANDEL, EINZELHANDEL)
041	2	TAETIGE INHABER (GROSSHANDEL, EINZELHANDEL)
042	2	VERKAUFSAUFSICHTSKRAEFTE UND EINKAEUFER
043	2	TECHNISCHE VERKAEUFER, HANDELSREISENDE UND HANDELS-
		VERTRETER
044	2	VERSICHERUNGSVERTRETER, VERSICHERUNGS-, IMMOBILIEN-
		UND BOERSENMAKLER, VERMITTLER GESCHAEFTLICHER
		DIENSTLEISTUNGEN UND VERSTEIGERER

segment

045	1	VERKAEUFER, VERKAUFSHILFSKRAEFTE UND VERWANDTE BERUFE
049	1	VERKAUFSHILFSKRAEFTE, SOWEIT NICHT ANDERWEITIG KLAS-SIFIZIERT
050	1	GESCHAEFTSFUEHRER IN GASTSTAETTEN- UND BEHERBERGUNGS-UNTERNEHMEN
051	1	TAETIGE INHABER VON GASTSTAETTEN UNE BEHERBERGUNGSUN-TERNEHMEN
052	1	HAUSWIRTSCHAFTLICHE UND VERWANDTE AUFSICHTSKRAEFTE
053	1	KOECHE, KELLNER, BARMIXER UND VERWANDTE BERUFE
054	1	HAUSGEHILFINNEN UND VERWANDTE HAUSWIRTSCHAFTLICHE BERUFE, SOWEIT NICHT ANDERWEITIG KLASSIFIZIERT
055	1	GEBAEUDEMEISTER, RAUM-, GEBAEUDEREINIGER UND VER-WANDTE BERUFE
056	1	WAESCHER, CHEMISCHREINIGER, BUEGLER
057	1	FRISEURE, SCHOENHEITSPFLEGER UND VERWANDTE BERUFE
058	1	SICHERHEITSBEDIENSTETE
059	1	DIENSTLEISTUNGSBERUFE, SOWEIT NICHT ANDERWEITIG KLASSIFIZIERT
060	2	LANDWIRTSCHAFTLICHE VERWALTER UND GUTSAUFSEHER
061	2	LANDWIRTE (EINSCHLIESSLICH SPEZIALISIERTE LANDWIRTE)
062	1	LAND- UND TIERWIRTSCHAFTLICHE ARBEITSKRAEFTE
063	1	FORSTARBEITSKRAEFTE
064	1	FISCHER, JAEGER UND VERWANDTE BERUFE
070	2	AUFSICHTSKRAEFTE DER PRODUKTION UND ALLGEMEINE VORMAENNER
071	2	BERGLEUTE, STEINBRECHER, TIEFBOHRER UND VERWANDTE BERUFE
072	2	HUETTENWERKER, GIESSER, HAERTER UND VERWANDTE BERUFE
073	1	HOLZAUFBEREITER, PAPIERHERSTELLER
074	2	CHEMIEWERKER UND VERWANDTE BERUFE
075	1	SPINNER, WEBER, STRICKER, FAERBER UND VERWANDTE BERUFE
076	1	GERBER, FELLZURICHTER, RAUCHWARENZURICHTER
077	1	NAHRUNGSMITTEL- UND GETRAENKEHERSTELLER
078	1	TABAKAUFBEREITER, TABAKWARENHERSTELLER
079	1	SCHNEIDER, DAMENSCHNEIDERINNEN, NAEHER, POLSTERER UND VERWANDTE BERUFE
080	1	SCHUHMACHER, LEDERWARENMACHER
081	1	MOEBELTISCHLER UND VERWANDTE HOLZBEARBEITER
082	1	STEINBEARBEITER, STEINBILDHAUER
083	2	GROBSCHMIEDE, WERKZEUGMACHER, WERKZEUGMASCHINENBE-DIENER
084	2	MASCHINENSCHLOSSER, MASCHINENMONTEURE UND PRAEZI-SIONSINSTRUMENTENMACHER (AUSGENOMMEN FUER ELEKTRISCHE MASCHINEN UND GERAETE)
085	2	ELEKTROMECHANIKER UND VERWANDTE ELEKTRO- UND ELEKTRO-NIKWERKER
086	1	SENDESTATIONSBEDIENER, TONAUFNAHME- TONWIEDERGABEAN-LAGENBEDIENER UND FILMVORFUEHRER
087	1	ROHRINSTALLATEURE, SCHWEISSER, BLECH- UND BAUMETALL-VERFORMER UND METALLBAUMONTEURE
088	1	SCHMUCKWARENHERSTELLER, EDELMETALLBEARBEITER
089	1	GLASVERFORMER, TOEPFER UND VERWANDTE BERUFE
090	1	GUMMI- UND KUNSTSTOFFWARENMACHER
091	1	PAPIERWARENMACHER, KARTONAGEMACHER
092	1	DRUCKER UND VERWANDTE BERUFE
093	1	MALER
094	1	GUETERERZEUGENDE UND AEHNLICHE BERUFSTAETIGKEITEN, SOWEIT NICHT ANDERWEITIG KLASSIFIZIERT
095	1	MAURER, ZIMMERER UND ANDERE BAUARBEITER
096	1	BEDIENER (MASCHINISTEN) STATIONAERER (KRAFT-)MASCHI-NEN UND AEHNLICHER ANLAGEN
097	1	BEDIENER VON MATERIALBEWEGUNGSGERAETEN UND AEHNLICHEN EINRICHTUNGEN; HAFEN- UND LADEARBEITER
098	1	TRANSPORTEINRICHTUNGSBEDIENER
099	1	HANDLANGER, UNGELERNTE HANDARBEITER ("LABOURES"), SOWEIT NICHT ANDERWEITIG KLASSIFIZIERT
101	9	SOLDAT (WEHRBERUF)
102	9	OFFIZIER (WEHRBERUF)
103	9	ARBEITSSUCHENDE NEUE ARBEITSKAREFTE
104	9	ARBEITSKRAEFTE MIT NICHT BESTIMMBAREN ODER UNZULAENG-LICH BESCHRIEBENEN BERUFEN
105	9	IN AUSBILDUNG
106	9	NICHT ERWERBSTAETIGE HAUSFRAUEN
107	9	RENTNER UND PENSIONAERE OHNE FRUEHERE BERUFSANGABE
108	9	KEINE BERUFSANGABE
109	9	WN
000	9	TNZ

Appendix B

Mapping of the industry response codes into
the industry-sectors

This classification follows the one from Beck et al. (1978;
see also Diekmann 1985: chapter 8).

industry	sector
agriculture,forestry	periphery
energy,water,mining	core
chemicals,stones,earth	core
iron,steel,other metals	core
machinery,cars,steel construction	core
office machines	core
electrotechnical,mechanics,optics	core
wood,paper,leather,textiles	periphery
food	periphery
construction	core
retail	periphery
transportation,communications	core
banking,insurances	core
other private services	periphery
churches,associations	periphery
public service	public service

References

Allison,P.D.(1985) Survival Analysis of Backward Recurrence Times. JASA 80: 315-322.

Baron,J.N. und W.T. Bielby(1980) Bringing the Firms Back in: Stratification, Segmentation, and the Organization of Work. ASR 45: 737-765.

Baron,J.N. und W.T. Bielby(1984) The Organization of Work in a Segmented Economy. ASR 49: 454-473.

Beck,E.M.; Horan,P.M. und C.M. Tolbert II(1978) Stratification in a Dual Economy: A Sectoral Model of Earnings Determination. ASR 43: 704-720.

Bielby,W.T. und J.N. Baron(1983) Organizations, Technology and Worker Attachment to the Firm. RSSM 2: 77-113.

Brüderl,J.(1986) Organisationale Determinanten von Berufskarrieren.
München (mimeo).

Carroll,G.R. und K.U. Mayer(1986) Job-Shift Patterns in the Federal Republic of Germany: The Effects of Social Class, Industrial Sector, and Organizational Size. ASR 51: 323-341.

Dickens,W.T. und K. Lang(1985) A Test of Dual Labor Market Theory. American Economic Review 75: 792-805.

Diekmann,A.(1985) Einkommensunterschiede zwischen Frauen und Männern.
Wien, Institut für Höhere Studien (mimeo).

Doeringer,P.B. und M.J. Piore(1971) Internal Labor Markets and Manpower Analysis.
Lexington:Heath.

Freeman,R.B.(1980) The Exit-Voice Tradeoff in the Labor Market: Unionism, Job Tenure, Quits, and Separations. Quarterly Journal of Economics 94: 643-673.

Hodson,R.(1983) Workers' Earnings and Corporate Economic

Structure.

New York:Academic Press.

Hodson,R.(1984) Companies, Industries, and the Measurement
of Economic Segmentation. ASR 49: 335-348.

Osterman,P.(1975) An Empirical Study of Labor Market
Segmentation.

Industrial and Labor Relations Review 28: 508-523.

Pfeffer,J.(1977) Toward an Examination of Stratification in
Organizations. Administrative Sc. Quart. 22: 553-567.

Rosenbaum,J.E.(1984) Career Mobility in a Corporate Hierarchy.
Orlando:Academic Press.

Sørensen,A.B.(1977) Estimating Rates from Retrospective
Questions, in: Heise,D.R.(ed.) Sociological Methodology
1977, 209-223.

San Francisco:Jossey-Bass.

Sørensen,A.B.(1983) Sociological Research on the Labor
Market.

Work and Occupations 10: 261-287.

Spaeth,J.L.(1984) Structural Contexts and the Stratifi-
cation of Work. RSSM 3: 77-92.

Stolzenberg,R.M.(1978) Bringing the Boss Back in: Employer
Size, Employee Schooling, and Socioeconomic Achievement.
ASR 43: 813-828.

Tolbert,C.II(1982) Industrial Segmentation and Men's Career
Mobility. ASR 47: 457-477.

Tuma,N. und M.T. Hannan(1984) Social Dynamics.
Orlando:Academic Press.

Wallace,M. und A.L. Kalleberg(1981) Economic Organization
of Firms and Labor Market Consequences: Toward a Speci-
fication of Dual Economy Theory, in: Berg,I.(ed.)
Sociological Perspectives on Labor Markets, 77-117.

New York: Academic Press.

Abbreviations:

ASR   American Sociological Review
JASA   Journal of the American Statistical Association
RSSM   Research in Social Stratification and Mobility

Departures from an internal labor market

Trond Petersen and Seymour Spilerman

1. Introduction

When do employees leave their organizations? For what reasons do they leave? An extensive literature on turnover tries to answer these questions (for a review see Bluedorn, 1982).

The literature on turnover is complemented by a parallell and almost independent literature on careers within organizations (see Rosenbaum, 1979, 1984; Wise, 1975; Halaby, 1982; White and Althauser, 1984; Skvoretz, 1984; Stewman and Konda, 1983). This second literature deals with questions about who gets promoted, after how much time, and how the position in the organization affects the rate of promotion.

A third, and more theoretical literature, gives ample reasons for why the decision to leave is not independent of the career prospects within the organization, and hence for why quit behavior should not be studied independently of organizational careers. For example, the literature on internal labor markets argues that employers sometimes implement firm internal promotion schemes in order to reduce turnover (see the recent review in Osterman, 1984). Yet, in the empirical literature promotions and departures have

been treated as almost independent events. Exeptions
are a few empirical studies that address both firm
internal and external mobility, but where the focus is
not on how they are related (see Sorensen and Tuma,
1981; Felmlee, 1982).

The intent of the paper is mainly empirical, but
the findings do have some broader conceptual
implications. They point to the need for an integrated
approach to both turnover and promotions. We need to
address how organizational constraints on intrafirm
careers shape departure decisions.

The paper is organized as follows. In the next
section we specify the main conceptual ideas and
derive hypotheses for the empirical investigation. In
section 3 the data, the personell records of a large
United States insurance company, are described.
Section 4 discusses the statistical method, a multi-
state hazard rate model. In sections 5 and 6 the
results are presented.

2. Departures and Careers: Hypotheses for research

Employees leave organizations for a variety of
reasons. We consider two reasons that are the outcomes

of choices employees make: leaving an organization for
career reasons and leaving it for personal reasons.
Career reasons are defined as better opportunities,
higher earnings, better working conditions, and more
interesting and suitable work, in jobs outside the
organization. Personal reasons are tending family
needs, such as illness. In both cases the employee
makes the decision to move, in contrast to cases where
he or she gets fired or laid off. The latter reasons
are dealt with in a separate paper (Petersen and
Spilerman, 1986a).

In studying the departure rates for these two
reasons we have three objectives. Our first objective
is to study how demographic groups differ in their
reasons for departure. Which groups leave for career
reasons and which for personal? How do men and women
differ in their reasons for departure?

Our second objcetive is to study how the rates of
leaving for the two reasons depend on the employee's
position in an organization. Are employees higher up
in an organizational hierarchy less likely to leave
than employees lower down?

Our third objective is to assess how the departure
rates are related to promotion opportunities in an

organization. Are departure rates low in locations where promotion rates are high? We want to assess the extent to which internal labor markets and promotion rates reduce turnover, as suggested by theories of internal labor markets.

In order to address the three objectives we shall be a bit more formal. Consider an employee who at each instant of calendar time τ is faced with choosing between three options: leaving the organization for career reasons, leaving it for personal reasons, or remaining in the organization with possibilities of future promotions as well as demotions. Call the three options C, P and R.

Let $E(C(\tau))$, $E(P(\tau))$ and $E(R(\tau))$ denote the expected future values of pursuing a specific option. The values will depend on monetary factors, such as future income gains and losses, and on psychological factors, such as satisfaction with the activity implied by the chosen option.

The consequences of the choice the employee makes will always be uncertain. For example, characteristics of the next job are not known with certainty, nor is there certainty with respect to future promotions, dismissals and demotions. Therefore, the value of

making a certain choice can only be assessed in terms of its expectation.

It seems reasonable to assume that the employee will prefer options with higher expected values to those with lower. Thus if $E(C(\tau))$ is greater than both $E(R(\tau))$ and $E(P(\tau))$ the employee will leave for career reasons.

How are the expected values determined? Consider first the expected value of leaving for personal reasons. Clearly, this depends on the employee's sex. Taking care of family is an activity that society values highly for women, but not so highly for men. Men receive little prestige for leaving their jobs in order to tend to their families. Women more often than men, therefore, may rank the strategy of leaving for personal reasons above the two other strategies. This translates into an hypothesis about the effects of demographic characteristics:

Hypothesis 1: Women leave organizations for personal reasons more often than men, keeping other things constant.

Consider next the expected value of leaving an organization for career reasons. That value depends in

large measure on the alternatives elsewhere available
as compared to the prospects in the organization. If
the employee's rewards in the organization already are
high, opportunities outside the company are likely to
be less attractive. We can state the second
hypothesis, which relates to the employee's positions
within the organization:

Hypothesis 2: The higher the current rewards the less
likely it is that the employee will leave the company
for personal or career reasons, keeping other things
constant.

In operational terms this means that for a given
rate of promotion, the higher the current rewards the
less likely it is that a departure will occur.

Consider finally the expected value of remaining in
the organization. This value will, for a given level
of already obtained rewards, clearly depend on the
probabilities of getting promoted in the future. In
locations in the company where promotion opportunities
are poor the expected value of remaining in the
organization is low. Thus, our third hypothesis becomes:

Hypothesis 3: Where promotion probabilities are low
departure rates are high, keeping other things
constant.

3. Data and Variables

The data used in this study are taken from the
personell records of a large United States insurance
company to which the authors have access. The company
employs approximately 16000 individuals. We use the
personell records pertaining to the career experiences
of every employee who either was in the company as of
1970 or who entered later but before the end of the
study in December 1978. Detailed information is
available about the timing (year, month and day) of
promotions, demotions and departures. Employees who
voluntarily left the company were asked to state the
main reason for doing so. Altogether 19 reasons are
recorded.

The company is hierarchically organized in salary
grade levels, from grade 1 (the lowest) to grade 20
(the highest). The hierarchy is explicit in written

documents and is clearly perceived by the employees.
Internal labor markets are well-developed. Vacant
positions are posted within the company and employees
are encouraged to apply for these. Positions are made
accessable to jobseekers outside the company only when
no suitable internal replacement can be found.

We analyse the rate of leaving a grade level, for
each of the three reasons discussed in the preceding
section. We do this by means of survival analysis (see
Tuma and Hannan, 1984). The dependent duration
variable in the analysis is the number of months an
employee spends in a _salary_ _grade_ _level_ before a
promotion, demotion, departure from the organization
or censoring (end of study, December 1978) occurs. A
salary grade level can be left for several reasons. In
our general formulation we consider 15 mutually
exclusive and exhausting reasons. In the present paper
we report only the analysis for the three reasons
discussed earlier: the grade level was left (1)
because the employee left the company for reasons tied
to his or her career; (2) because he or she left the
company for personal reasons; (3) because he or she
got promoted to a higher grade level. We deal with
dismissals and demotions (quite rare) in a separate

paper (Petersen and Spilerman, 1986a). Definitions of
the three reasons are given in Table 1.

(Table 1 about here)

The rates of leaving a grade level for any of the
three reasons, career, personal, or promotion, are
predicted using three sets of covariates: demographic,
human capital and organizational. The demographic
variables are race and sex. The human capital
variables are educational level (from level 0 to 9),
seniority in the company and the employee's age at the
time the current grade level was entered. In addition,
the rate of leaving a grade level at duration t in the
grade also depends on t; thus allowing for duration
dependence, positive or negative. The organizational
variables are the division of the company in which the
employee works, his or her job focus (i.e. a company-
specific occupational code) and location (home office
versus other). In the present analysis job focus and
division are included mainly as control variables. The
job focus variable is however of substantive interest
in its own right and is dealt with in separate papers
(see Spilerman and Petersen, 1986; Petersen and

Spilerman, 1986b). Further definitions of the
variables are given in Table 1.

In the statistical analysis we fix the seniority
and age variables at their values at the date a grade
level was entered, but allow them to change when a
change in grade level occurs. They depend on time
between grade levels, but not within grade levels. The
covariates division, job focus and location are
treated as time-dependent within as well as between
grade levels. If an employee changed his or her job
focus while remaining in the same grade level it is
taken into account in the analysis (for details see
Petersen, 1986a, 1986b).

By introducing both seniority in the company and
duration in the grade level we allow for the
possibility that duration in a grade may have a
different effect than seniority in the company, on,
say, the rate of getting promoted. The behavior of the
employee may be governed by two "clocks" - duration in
grade and seniority in company - with possibly
opposite effects.

The focus of the present analysis is on the
demographic, human capital, location and duration
effects on the three rates considered, leaving the

discussion of the effects of job focus and division to the papers cited above.

4. Methods

We specify a continuous time multi-state hazard rate model. Let the hazard rate for leaving salary grade level s for reason j after duration t in the grade be

$$(1) \quad \lambda_{sj}(t|x(t)) = \lim_{\Delta t \downarrow 0} P[t \leq T < t + \Delta t, J = j | T \geq t, x(t)]/\Delta t,$$

where T is a random variable denoting the duration in grade s and J another random variable denoting the reason for which the grade was left. x(t) is the set of covariates that influences the rate, evaluated at duration t and possibly summarizing the employee's history in the company up to duration t in grade s. The index s runs from 1 through 20, one for each of the salary grade levels, while j runs from 1 thorugh 15, one for each of the 15 reasons for leaving the grade.

The overall rate of leaving grade s is given as

$$(2) \quad \lambda_s(t|x(t)0 = \sum_{j=1}^{J} \lambda_{sj}(t|x(t)),$$

where J is equal to 15. Here we report only estimates
of λ_1, λ_2 and λ_3, the other reasons being dealt with
in separate papers. Also, estimates are reported only
for grades 2, 4, 7, and 12-15 (the latter four grouped
together). The points to be made come through by
considering this subset of grades, and the savings in
the number of tables to be reported is enormous.

From (1) and (2) it follows that the probability
that grade s was left for reason j, given that it was
left after duration t is

(3) $P_s[J=j|T=t,x(t)] = \lambda_{sj}(t|x(t))/\sum_{j=1}^{J}\lambda_{sj}(t|x(t))$.

Thus the parameters pertaining to the rates not only
tell us how long an employee waits before experiencing
a transition, but also the probability of a specific
type of transition, given that one occurred. Equation
(3) has the form of a multinomial logit model, and its
interpretation is well-known.

It should be stressed that the specification in (1)
or (2) makes no assumption about independence of the
different reasons (for discussions of this point see
Prentice et al., 1978: 545-547). An objective of this

analysis is to investigate how the three rates change
in the same or opposite directions as one moves up the
hierarchy of the company, and any independence
assumption would be illegitimate. See Hypothesis 3 of
section 2.

Each of the cause-specific rates is given a log-
logistic specification (see Kalbfleisch and Prentice,
1980:27-28). Without going into the details this
specification is particularily suitable in the present
context. If the effect of duration on, say, the rate
of getting promoted is positive, then the rate of
promotion is given as a bell-shaped function of
duration in the grade. During the initial months spent
in the grade the rate of experiencing a promotion
increases with time; then it reaches a peak;
whereafter it decreases with time. If the effect of
duration is negative the rate declines with time over
the entire time spent in the grade. We expect to find
a bell-shaped effect of duration on the rate of
getting promoted. Figure 1 illustrates the shapes the
log-logistic hazard allows.

(Figure 1 about here)

The interpretation of the parameters pertaining to
the other covariates is the usual: a variable with a
positive parameter increases the rate and one with a
negative decreases it. Or put this way, a positive
parameter translates into a shorter waiting time
before an event of the specific type occurs while a
negative translates into a longer waiting time.

The parameter estimates are obtained by the method
of maximum likelihood, as described in Petersen
(1986a).

5. Descriptive statistics

Table 2 gives descriptive statistics for the
variables used in the analysis, other than division
and job focus, which presently are not of substantive
interest. When appropriate, means and standard
deviations are reported, otherwise the proportions of
employees having a certain value on a variable are
given.

(Table 2 about here)

Table 2 shows that the four variables - education,

duration in grade level, seniority and age (as measured at the date the grade was entered) - all increase from one grade level to the next. Employees higher up in the hierarchy have been longer in the company, are older and are better educated.

The proportion of white and the proportion of male employees increase as we move up the hierarchy. The company's home location has a higher concentration of high grade levels than the other locations.

In Table 2 five destination states are considered: the employee was still in the grade level at the time the study ended (a censored observation), he or she left the company due to one of three reasons (career, personal or other), and finally he or she got promoted from the grade. In the analysis of rates we focus only on promotions, career and personal reasons for leaving the company.

We see that the proportions who left the company, for any of the three reasons, decline sharply with the grade level. This may in part reflect the fact that the employees in the higher grade levels are older and therefore less likely to leave. In part it may be the case that the higher grade levels offer their employees more opportunities. The alternatives

elsewhere available may be less attractive to those higher up in the hierarchy.

Further, we see that the proportions being promoted from a grade increase with the grade level. This may reflect in part higher promotion rates in the higher grade levels, but also the fact that the departure rates are lower higher up, so that more employees remain in the company until a promotion occurs.

6. Analysis of rates

We report the results on rates in two parts. First, we discuss the effects of the independent variables on the three rates, for each grade level considered. Then we discuss how the rates depend on the grade level.

Effects of independent variables

The estimates of the three first rates of the multi-state hazard in equation (2) are given in the Appendix, Tables 5-8, one table for each set of grades, 2, 4, 7 and 12-15. In each table three different specifications of the rates are considered. In Panels A only the demographic variables enter, i.e.

race and sex, and of course duration. In Panels B the human capital variables - education, age and seniority - are added to the variables in Panels A. In Panels C we finally add the organizational variables.

Let me first point out that the models in Panels C improve the fit significantly compared to the models in Panels B, which in turn improve the fit significantly relative to the models in Panels A, using a likelihood ratio test statistics.[1]

Tables 5-8 contain a great deal of information. Here we report only the most salient points, leaving more detailed analysis to other papers. Table 3 gives a summary of the main findings. A short explanation of this table is in order. In the top row of Table 3 we list the four set of grade levels considered. For each grade there are three hazard rates. The far left column lists the variables whose effects we report. The entries in the interior of the table give the effects of particular variables on each of the rates in each grade level. An entry of two plus signs (++) means that the designated rate increases strongly with an increase in the variable; one plus sign means that it increases; a zero sign (0) means that the variable has little or no effect; a negative sign (-) means

that it has a negative effect and two negative signs
(--) means that the effect is strongly negative. The
sign ? means that the effect is ambiguous, and applies
often to the effects of education. The sign O- (or O+)
means that while the effect is negative (or positive)

(Table 3 about here)

Six main findings appear from Table 3. First, in
all grades women have lower rates than men of leaving
the company for career reasons, but much higher rates
of leaving for personal reasons. This reflects no
doubt the different roles of men and in the labor
force. Women often leave jobs to take care of their
families or because their husband took a job somewhere
else, whereas men rarely leave for those reasons, as
predicted by Hypothesis 1. In the lower grades, 2, 4,
and 7, women have lower rates of promotion than men,
while in the higher grades, 12-15, they are at an
advantage relative to men.

Second, in all grades blacks have the lowest rates
of departures and promotions, particularily in grades
2, 4 and 7. This means that they remain longer in each
grade level. In part they do so because they wait

longer before getting promoted, and in part because they wait longer before quitting the company. Their lower rates of promotion probably reflect the disadvantage of being black in general, not of being black in this particular company, otherwise one would have expected higher quit rates for blacks. They would have sought jobs in other companies with more favorable treatment of blacks, had such companies existed. This point can only be appreciated by considering the multi-state hazard estimated in this paper, and not from the hazard for promotion alone.

Third, age and seniority in the company have negative effects on all three rates in all grade levels. The older a person is and the longer he or she has been in the company the less likely he or she is to get promoted or to leave. The effect of seniority is stronger than the effect of age; an increase in seniority of one year has a stronger negative effect on the rate than the same increase in age has. This holds in almost all of the rates estimated.

Fourth, the effects of duration are interesting. In all grades the effect of duration on the rate of promotion is positive. This means that the probability of experiencing a promotion in, say, the next month

increases with duration in the grade, up to some
point, whereafter it decreases. The hazard has a bell-
shaped function as illustrated in Figure 1. On the two
other rates the effect of duration is ambiguous, and
varies with the grade. In grade 2, for example,
duration has a negative effect on the rate of leaving
for career reasons, while it has a positive effect on
the rate of leaving for personal reasons. Both
effects, however, are quite small. Figure 2
illustrates the shapes of the the three hazards as
functions of duration, for grade 2.

(Figure 2 about here)

Fifth, we see that location has a positive effect
on the rate of promotion in grades 2 and 4, while
negligible effects in grades 7 and 12-15. In grades 2
and 4 the effects of location on the two departure
rates are negative. The interpretation of this finding
seems simple. In the lower grade level in the home
office opportunities for promotions are high, and
departure rates are consequently low. The results
indicate that departure rates are not independent of
the promotion rates, and that consideration of the

multi-state hazard yields valuable insights. The
finding confirms the prediction of hypothesis 3, see
section 2.

Sixth, the effects of education are ambiguous.
Education has a clear effect on the rate of promotion
in all grades but grade 2. The more education the more
likely one is to be promoted. For the other rates no
similar inference can be drawn.

Differences in rates between grade levels

In the preceeding section the effects of covariates
were assessed, keeping the grade level constant. Now
we turn to an analysis of the effect of the grade
level itself, keeping covariates constant.

This analysis can be accomplished in a variety of
ways, depending on which covariates one chooses to
keep constant. We focus on how the three rates vary
with the grade level for two groups of employees:
white males and white females. Since the rates exhibit
duration dependence we also need to choose values of
duration in grades for which we want to assess the
rates. We estimate the rates after 1 month duration in
the grade levels.

The estimates of the rates are computed from Panels
A of Tables 5-8. These rates should be considered as
average rates for white men and for white women, where
we average over education, age, seniority etc., while
keeping race constant.

Table 4 gives the estimated rates for each grade
aand sex.

(Table 4 about here)

Two things are striking in Table 4. First, we see
that the rate of getting promoted is more or less the
same across the grades. This suggests that pyramidal
organizational structures need not constrain careers
in the higher echelons of the hierarchy more than
careers in the lower. Employees in the upper echelons
of the hierarchy are as likely to be promoted as those
lower down (see also Stewman and Konda, 1983). The
important factor is the ratio between jobs at grade k
and jobs at grade k+1. This ratio need not decline
with k. This issue is pursued in detail in our other
papers.

Second, we see that the rates of leaving the

company, for both of the two reasons considered,
decline sharply with the grade level. In particular,
the rate of leaving for personal reasons becomes very
close to zero in the highest grade level. As shown in
Table 2, only half a percent of those in grades 12-15
left for personal reasons, while as much as 11 percent
in grade 2 did. This confirms the prediction of
Hypothesis 2, that keeping promotion rates constant,
the departure rate will decline with the level of
already obtained achievement in the corporation.

One may speculate on the reasons for the second
finding. One plausible explanation is that in the
higher grade levels employees have accumulated firm
specific human capital, and their opportunities
outside the company are therefore more restricted than
their opportunities inside. The relative value of
staying in the company is higher than for those in
lower grade levels, in the sense that the alternatives
available outside are not as attractive as those
available inside.

In Figure 3 these effects of the grade levels are
graphed. Along the horizontal axis the grade levels
are drawn, while   the vertical shows the rates. The
Figure displays clearly how the rates of departure

decline while the rates of promotion are more or less
constant.

(Figure 3 about here)

7. Conclusions

In this paper we have studied the determinants of
voluntary departures for career and personal reasons
from a large internal labor market. We assessed how
quit rates were related to promotion rates in the
organization, to the employee's position in the
organization and to demographic and human capital
characteristics. The company studied is hierarchically
organized in salary grade levels. Using a multi-state
hazard rate model we predicted the rates of getting
promoted from a grade level and the rates of leaving
it for either career or personal reasons.

Several findings were reported. Here we summarize
the most important. First, it was shown that women
have lower rates of leaving for career reasons than
men, but much higher rates of leaving for personal
reasons. This reflects the basic social difference
between men and women in dealing with family obligations.

We considered simultaneously the effects of age, seniority and duration in grade level on the three rates. Age and seniority, measured relative to the date a grade level was entered, had negative effects on all three rates in all grade levels. Duration in grade level, in contrast, had a positive effect on the rate of getting promoted. The probability of experiencing a promotion increases with time during the initial months in a grade level, whereafter it peaks and then declines. The effects of duration on the two rates for departure were ambiguous, either not very strong or not significant.

Second, it was shown that the rates of departure are lower in the higher echelons of the organization, whereas the rates of promotion remain more or less constant as one moves up the career ladder. Thus, employees who have reached positions of high achievement are less likely to leave than those lower down in the hierarchy.

Third, it was found that in structural positions where rates of promotions are high, i.e. in the organziation's home office in grades 2 and 4, the rates of departure are low. This suggests that the

process governing leaving should not be studied

independently of the process governing internal

promotions. To understand the first process we also

need to understand the second.

Bibliography

Baron, James N. 1984. "Organizational perspectives on stratification." Annual Review of Sociology 10: 37-69.

Becker, Gary S. 1975. Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education (second edition). Chicago: Chicago University Press.

Bluedorn, Allen C. 1982. "The theories of turnover: Causes, effects, and meaning." Pp. 75-128 in Samuel B. Bacharach (ed.), Research on the Sociology of Organizations. JAI Press.

Burawoy, Michael. 1979. Manufacturing Consent. Chicago: Chicago University Press.

Carter, Michael J., and Susan B. Carter. 1985. "Internal labor markets in retailing: The early years." Industrial and Labor Relations Review 38(4):586-598.

Felmlee, Diane H. 1982. "Women's job mobility processes." American Sociological Review 47:142-151.

Jacobs, David. 1981. "Towards a theory of mobility and behavior in organizations: An inquiry into the

consequences of some relationships between individual performance and organizational success." American Journal of Sociology 87 (3): 684-707.

Halaby, Charles N. 1979. "Sexual inequality in the work place: An employer-specific analysis of pay differences." Social Science Research 8:79-104

Halaby, Charles N. 1980. "Dynamic models of attainment in the workplace." Social Science Research 9(1): 1-36..

Halaby, Charles N. 1982. "Job-shift differences between men and women in the workplace." Social Science Research 11(1): 1-29.

Kalbfleisch, J.D., and R.L. Prentice. 1980. The Statistical Analysis of Failure Time Data. New York: Wiley.

Kalleberg, Arne L., and Aage B. Sorensen. 1979. "The Sociology of Labor Markets." Annual Review of Sociology 5: 351-379.

Osterman, Paul. 1984. "Introduction: The nature and importance of internal labor markets." Pp. 1-22 (ch. 1) in Paul Osterman (ed.), Internal Labor Markets. Cambridge, MA: MIT Press.

Petersen, Trond .1986a. "Estimating fully

parametric hazard rate models with time-dependent covariates. Use of maximum likelihood." Sociological Methods and Research 14(3):219-246.

Petersen, Trond .1986b. "Fitting parametric survival models." To appear in Journal of the Royal Statistical Society.

Petersen, Trond, and Seymour Spilerman. 1986a. "Involuntary departures from an internal labor market." Manuscript in progress.

Petersen, Trond, and Seymour Spilerman. 1986b. "The structure of an internal labor market and departures." Manuscript in progress.

Pfeffer, J. 1977. "Toward an examination of stratification in organizations." Administrative Science Quarterly 22(4):553-567.

Prentice, R.L., J.D. Kalbfleisch, A.V. Peterson, jr., N. Flournoy, V.T. Farewell, and N.E. Breslow. 1978. "The analysis of failure times in the presence of competing risks." Biometrics 34:541-554.

Romanov, Allyn L. 1984. "A Brownian motion model for decision making." Journal of Mathematical Sociology 10(1):1-28.

1144-1175.

Rosenbaum, James E. 1979. "Organizational career mobility: Promotion chances in a corporation during periods of growth and contraction." American Journal of Sociology 85(1):21-48.

Rosenbaum, James E. 1984. Career Mobility in a Corporate Hierarchy. New York: Academic Press.

Sandefur, G. 1981. "Organizational boundaries and upward job shifts." Social Science Research 10:67-82.

Skvoretz, John. "Career mobility as a Poisson process. An application to the career dynamics of men and women in the U.S. Office of the COmptroller of the Currency from the Civil War to World War II." Social Science Research 13(2):198-220.

Sorensen, Aage B. 1983. "Sociological Research on the labor market. Conceptual and methodological issues." Work and Occupations 10(3):261-287.

Sorensen, Aage B., and Arne L. Kalleberg. 1981. "An outline of a theory of the matching of persons to jobs." Pp. 49-74 (ch. 3) in Ivar Berg (ed.), Sociological Perspectives on Labor Markets. New York: Academic Press.

Sorensen, Aage B., and Nancy B. Tuma. 1981. "Labor market structures and job mobility." Research on

Stratification 1:67-94.

Spence, A. Michael. 1974. Market Signalling.
Informational Transfer in Hiring and Related Screening
Processes. Cambridge, Mass.: Harvard University Press.

Spilerman, Seymour. 1986. "Organizational rules and
features of work careers." Forthcoming in Reseacrh on
Organizations.

Spilerman, Seymour, and Trond Petersen. 1986.
"Organizational structure and the determinants of
promotions." Manuscript, Columbia University and
Harvard University.

Stewman, Shelby, and Shuresh L. Konda. 1983.
"Careers and organizational labor markets: Demographic
models of organizational behavior." American Journal
of Sociology 88(4):637-685.

Stinchcombe, Arthur L. 1959. "Bureaucratic and
craft administration of production." Administrative
Science Quarterly 4 ():546-558.

Stinchcombe, Arthur L. 1974. Creating Efficient
Industrial Administrations. New York: Academic Press.

Tuma, Nancy B., and Michael T. Hannan. 1984. Social
Dynamics. Models and Methods. New York: Academic
Press.

Waters, L.K. and D. Roach. 1979. "Job satisfaction, behavioral intention, and absenteeism as predictors of turnover." Personell Psychology 32:393-397.

Weber, Max. 1978. Economy and Society. Berkeley: University of California Press.

White, Harrison C. 1970. Chains of Opportunity. Cambridge, Mass.: Harvard University Press.

White, Robert W., and Robert P. Althauser. 1984. "Internal labor markets, promotions, and workers skill: An indirect test of skill ILMs." Social Science Research 13(4):373-392.

Wise, David A. 1975. "Personal attributes, job performance, and probability of promotion." Econometrica 43(5-6):913-931.

NOTES

1. The value of the chi-square statistics is computed as follows. Let L_1 be the loglikelihood for a hazard with the added variables and L_0 for the hazard with those variables excluded. Minus twice the difference between L_0 and L_1 yields the value of the chi-square statistics. The degrees of freedom is equal to the number of parameters added between the model with L_1 and the model with L_0.

Table 1

Definition of Variables Used in the Analysis

Duration in grade: Measured as months spent in current salary grade level.

Age: Measured in years as of starting date of current salary grade level.

Seniority: Measured as the number of months the employee has spent in the company as of the starting date of the current salary grade level.

Definition of destination states:

State 1: Left for career reasons:

    Higher earnings
    Better working conditions
    Greater opportunity
    More interesting or suitable work

State 2: Left for personal reasons:

    Nearer home or better transportation
    Change of residence
    Household duties
    Illness in family

State 3: Got promoted within company:

    Any move leading to a salary grade level higher than
    the grade currently occupied

Educational level:

0	Less than four years of high school
1	High school graduate (4 years)
2	High school graduate (4 years) plus secretarial or business school
3	College courses or certificates, less than 60 credits
4	College courses, 60 or more credits but degree not received
5	Junior or community college degree
6	Bachelor's degree
7	Graduate school courses, advanced degree not received
8	Master's degree
9	Doctorate

Division:

1	Agency
2	Corporate
3	Group
4	Individual
5	Investment

Job Focus:

1	Machine-operator
2	Secretary/steno
3	Typist
4	Figure Clerk
5	Other Clerk
6	Accounting Claims/Contract Analysis
7	Math-programming
8	Sales-Staff
9	Underwriting-Investment
10	Not Applicable
11	Other Codes

Location:

1 = if employee works in home office
0 = otherwise

Table 2

Descriptive Statistics for Variables Used in the Analysis
By Salary Grade Level (Excluding Division and Job Focus)

	Salary Grade Level			
Variables	2	4	7	12-15

Means and Standard Deviations:

Duration in months[a]	14 (14)	19 (17)	22 (18)	22 (18)
Seniority in months[b]	6 (11)	27 (39)	83 (91)	150 (120)
Age in years[b]	24 (8)	28 (10)	31 (10)	37 (9)
Educational level[c]	1.7 (1.3)	2.6 (1.9)	3.2 (2.3)	5.2 (2.3)

Proportions:

Race
White	.67	.73	.80	.94
Black	.23	.18	.13	.04
Oriental	.02	.02	.02	.01
Spanish	.08	.06	.05	.01

Sex
Female	.90	.90	.66	.20
Male	.10	.10	.34	.80

Educational Level
0 or 1	.61	.46	.38	.14
2	.13	.12	.08	.02
3	.16	.14	.15	.11
4 or 5	.07	.11	.10	.08
6, 7, 8 or 9	.03	.16	.29	.66

Location[d]
Home	.33	.33	.64	.64
Other	.67	.67	.34	.34

Destination States:
Remained in grade	.04	.16	.26	.29
Left the company	.46	.35	.15	.09
career reasons	.13	.12	.05	.03
personal reasons	.11	.09	.03	.005
other reasons	.22	.14	.07	.05
Promoted from grade	.49	.48	.58	.60

N (employees)	8894	8122	3235	4838

[a] The mean number of months spent in a grade level, including censored observations.

[b] Measured as of starting date of grade level.

[c] The mean of the educational score computed from the scale running from a low 0 to a high 9, see Table 1.

[d] Measured as of the date the grade either was left or censoring occurred.

Table 3

Summary of Effects of Independent Variables on the Three Rates By Grade level[a]

	Salary Grade Level											
	2			**4**			**7**			**12-15**		
	Rates[b]											
Variables	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
Duration	O-	O+	++	O+	O+	++	O-	+	++	+	O-	++
Black[c]	-	-	-	O	-	-	O+	O-	-	O-	O+	++
Oriental	O+	O-	+	O-	O+	O+	O-	O+	O-	O-	--	O-
Spanish	O-	O-	-	O-	O+	-	O+	O-	O-	O-	O+	O-
Female[d]	-	++	O-	-	++	-	O-	++	O-	-	++	+
Age	-	-	-	-	-	-	-	-	-	-		-
Seniority	-	-	-	-	-	-	-	-	-	-		-
Education[e]	?	+	?	?	?	+	?	?	++	?		+
Location	-	-	+	-	-	+	O-	O+	O-	O-		O-

[a] Computed from Panels C of Tables 5-8. The effects on λ_2 in grades 12-15 are computed from Panel A of Table 8, since too few transitions were made for personal reasons in grades 12-15 to estimate reliably the other effect parameters. See note e in Table 8 and the proportions and the N in Table 2.

[b] The rates λ_1, λ_2 and λ_3 are for respectively leaving the company for career reasons, leaving it for personal reasons, and getting promoted within the company. See Table 2 and section 3 for more precise descriptions of the three reasons.

[c] The race effects are measured relative to being white.

[d] The effect of being female is measured relative to being male.

[e] The entries summarize the effects of the educational dummy variables.

Legend: The entries mean:

++ the variable increases the rate strongly
+   the variable increases the rate
O+ the variable has a positive but either small or
    insignificant effect on the rate
O- the variable has a negative but either small or
    insignificant effect on the rate
-   the variable decreases the rate
-- the variable decreases the rate strongly
?   the variable has an ambiguous effect on the rate

Table 4

Estimated Rates After One Month Duration for Selected Values on Independent Variables By Grade level[a]

	Rates for white males			Rates for white females		
Grade Level	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
2	.02237	.00334	.00247	.01499	.01227	.00183
4	.01657	.00273	.00302	.01110	.00822	.00499
7	.00673	.00024	.00334	.00247	.00111	.00166
12-15	.00150	.00016	.00091	.00123	.00074	.00111

[a] Computed from Panels A of Tables 5-8.

[b] The rates λ_1, λ_2 and λ_3 are respectively for leaving the company (from the relevant grade level) for career reasons, leaving it for personal reasons, and for getting promoted within the company. See Table 2 and section 3 for more precise descriptions of the three reasons.

APPENDIX:

The estimates of the multi-state hazard in equation (2) follow in the next four tables, Tables 5-8. See section 6 for description of the results and Tables 3 and 4 for summaries of the main effects.

Table 5

Estimates of the Parameters of the Multistate Hazard Rate Model for Destination States 1, 2 and 3 from Salary Grade Level 2[a] (estimated standard errors in parentheses)

Panel A: Only Demographic Effects

	Destination State		
Independent Variables	Departure for Career Reasons	Departure for Personal Reasons	Promotion Within Company
Constant	-3.834 (.136)	-5.778 (.233)	-6.036 (.108)
Duration	-.074 (.030)	-.018 (.033)	1.227 (.027)
Race[b]			
Black	-.591 (.084)	-.573 (.091)	-.472 (.057)
Oriental	.153 (.273)	.084 (.309)	.568 (.206)
Spanish	-.500 (.124)	-.259 (.129)	-.265 (.084)
Female (=1)	-.330 (.108)	1.356 (.216)	-.294 (.081)
-Loglikelihood[c]	6415.1	5680.7	18014.1

Table 5   (Continued)

Panel B: Demographic and Human Capital Effects

	Destination State		
Independent Variables	Departure for Career Reasons	Departure for Personal Reasons	Promotion Within Company
Constant	-3.209 (.166)	-5.613 (.256)	-5.596 (.122)
Duration (in grade)	-.022 (.030)	.032 (.033)	1.284 (.028)
Race[b]			
Black	-.573 (.086)	-.513 (.092)	-.448 (.057)
Oriental	.155 (.278)	-.164 (.318)	.609 (.208)
Spanish	-.481 (.125)	-.206 (.129)	-.264 (.084)
Female (=1)	-.302 (.110)	1.357 (.217)	-.211 (.081)
Age (in years)	-.029 (.004)	-.009 (.004)	-.023 (.002)
Seniority (months)	-.040 (.004)	-.038 (.004)	-.016 (.002)
Educational Level[d]			
2	.365 (.095)	.086 (.106)	-.266 (.074)
3	.241 (.089)	.197 (.099)	.157 (.067)
4 or 5	.060 (.133)	.389 (.131)	.052 (.095)
6, 7, 8 or 9	.618 (.180)	.686 (.182)	-.041 (.158)
-Loglikelihood[c]	6319.9	5617.1	17919.5

Table 5   (Continued)

Panel C: Demographic, Human Capital and Organizational Effects

	Destination State		
Independent Variables	Departure for Career Reasons	Departure for Personal Reasons	Promotion Within Company
Constant	-2.101 (.220)	-4.976 (.297)	-6.521 (.159)
Duration (in grade)	-.001 (.031)	.044 (.034)	1.300 (.028)
Race[c]			
Black	-.290 (.090)	-.368 (.097)	-.559 (.061)
Oriental	.106 (.286)	-.115 (.323)	.527 (.207)
Spanish	-.232 (.130)	-.069 (.133)	-.338 (.087)
Female (=1)	-.875 (.123)	1.169 (.223)	-.133 (.081)
Age (in years)	-.035 (.004)	-.010 (.004)	-.021 (.002)
Seniority (months)	-.034 (.004)	-.036 (.004)	-.016 (.002)
Education Level[d]			
2	.049 (.100)	-.073 (.110)	-.266 (.074)
3	.069 (.093)	.086 (.103)	.157 (.067)
4 or 5	-.289 (.137)	.215 (.134)	.052 (.095)
6, 7, 8 or 9	.312 (.184)	.485 (.188)	-.041 (.158)
Job Focus[e]			
1	.328 (.123)	.168 (.125)	.290 (.088)
2	.135 (.122)	-.195 (.124)	.165 (.088)
3	.504 (.100)	.059 (.107)	.219 (.070)
4	.033 (.133)	-.054 (.141)	.459 (.094)
Division[f]			
2	-.751 (.196)	-.361 (.199)	.497 (.122)
3	-.560 (.122)	-.373 (.128)	.647 (.095)
4	-.532 (.119)	-.278 (.131)	.547 (.094)
5	-.320 (.196)	.169 (.199)	.565 (.149)
Location (1=Home)	-.959 (.103)	-.582 (.105)	.274 (.060)
-Loglikelihood[c]	6207.3	5588.7	17849.4

[a] The parameters of the multi-state hazard in equation (1) were estimated by the Method of Maximum likelihood, using the algorithm described in Petersen (1986a). Number of employees in grade level 2 is 8894.

[b] Excluded group: White and Native Americans.

[c] The loglikelihood is partioned into three parts, one for each of the three rates. The loglikelihood pieces for the full 15 state hazard are not reported in the Table. They are of no use in performing tests of interests for current purposes.

[d] Excluded group: High School Graduate or less education.

[e] Excluded group: Job foci with codes 6 and higher.

[f] Excluded group: Agency.

Table 6

Estimates of the Parameters of the Multistate Hazard Rate Model
for Destination States 1, 2 and 3 from Salary Grade Level 4[a]
(estimated standard errors in parentheses)

Panel A: Only Demographic Effects

	Destination State		
	Departure for Career Reasons	Departure for Personal Reasons	Promotion Within Company
Constant	-4.114 (.153)	-5.953 (.246)	-5.843 (.110)
Duration	-.131 (.033)	-.083 (.037)	.983 (.026)
Race[b]			
Black	-.366 (.099)	-.465 (.120)	.053 (.062)
Oriental	.247 (.223)	.410 (.257)	.250 (.193)
Spanish	-.847 (.188)	-.077 (.162)	-.170 (.099)
Female (=1)	-.342 (.114)	1.111 (.219)	-.549 (.080)
-Loglikelihood[c]	5614.0	4475.8	17437.4

Table 6   (Continued)

Panel B: Demographic and Human Capital Effects

	Destination State		
	Departure for Career Reasons	Departure for Personal Reasons	Promotion Within Company
Constant	-3.678 (.210)	-5.328 (.302)	-5.309 (.131)
Duration (in grade)	.022 (.034)	.027 (.038)	1.082 (.027)
Race[b]			
Black	-.228 (.104)	-.452 (.122)	.001 (.062)
Oriental	-.021 (.223)	.203 (.269)	.067 (.195)
Spanish	-.587 (.196)	-.007 (.166)	-.228 (.099)
Female (=1)	-.083 (.118)	1.265 (.223)	-.412 (.083)
Age (in years)	-.021 (.004)	-.026 (.005)	-.028 (.002)
Seniority (months)	-.044 (.002)	-.019 (.001)	-.005 (.001)
Educational Level[d]			
2	.327 (.124)	.282 (.127)	-.217 (.077)
3	.422 (.112)	.114 (.128)	.151 (.074)
4 or 5	.214 (.122)	-.078 (.147)	.153 (.083)
6	.407 (.106)	.369 (.120)	.067 (.080)
7, 8 or 9	.191 (.225)	-.252 (.309)	.252 (.180)
-Loglikelihood[c]	5244.4	4336.8	17273.5

Table 6 (Continued)

Panel C: Demographic, Human Capital and Organizational Effects

	Destination State		
	Departure for Career Reasons	Departure for Personal Reasons	Promotion Within Company
Constant	-2.724 (.286)	-4.457 (.367)	-6.173 (.187)
Duration (in grade)	.039 (.034)	.033 (.038)	1.103 (.027)
Race[c]			
Black	.011 (.107)	-.292 (.125)	-.142 (.065)
Oriental	-.040 (.233)	.227 (.278)	.024 (.195)
Spanish	-.366 (.202)	.164 (.175)	-.374 (.100)
Female (=1)	-.522 (.129)	1.057 (.228)	-.258 (.087)
Age (in years)	-.031 (.005)	-.031 (.005)	-.024 (.002)
Seniority (months)	-.038 (.002)	-.017 (.001)	-.005 (.001)
Educational Level[d]			
2	.077 (.128)	.139 (.131)	-.060 (.080)
3	.334 (.117)	.026 (.131)	.253 (.076)
4 or 5	.211 (.125)	-.137 (.151)	.229 (.085)
6	.433 (.115)	.284 (.132)	.187 (.087)
7, 8 or 9	.215 (.233)	-.331 (.316)	.332 (.184)
Job Focus[e]			
1	.005 (.266)	-.238 (.269)	-.316 (.141)
2	.396 (.171)	-.133 (.182)	-.093 (.115)
3	.302 (.217)	-.014 (.221)	-.251 (.130)
4	.146 (.211)	-.295 (.248)	-.237 (.139)
5	-.037 (.129)	-.198 (.146)	-.341 (.097)
Division[f]			
2	-.449 (.220)	-.727 (.246)	.848 (.116)
3	-.459 (.134)	-.384 (.143)	.732 (.090)
4	-.749 (.162)	-.290 (.152)	.648 (.094)
5	-.523 (.192)	-.899 (.250)	.650 (.094)
Location (1=Home)	-.737 (.119)	-.395 (.122)	.442 (.060)
-Loglikelihood[c]	5162.9	4310.3	17186.7

[a] The parameters of the multi-state hazard in equation (1) were estimated by the Method of Maximum likelihood, using the algorithm described in Petersen (1986). Number of employees in grade level 4 is 8122.

[b] Excluded group: White and Native Americans.

[c] The loglikelihood is partioned into three parts, one for each of the three rates. The loglikelihood pieces for the full 21 state hazard are not reported in the Table. They are of no use in performing tests of interests for current purposes.

[d] Excluded group: High School Graduate or less education.

[e] Excluded group: Job foci with codes 6 and higher.

[f] Ecluded group: Agency.

Table 7

Estimates of the Parameters of the Multistate Hazard Rate Model for Destination States 1, 2 and 3 from Salary Grade Level 7^a (estimated standard errors in parentheses)

Panel A: Only Demographic Effects

	Destination State		
	Departure for Career Reasons	Departure for Personal Reasons	Promotion Within Company
Constant	-5.026 (.285)	-8.320 (.491)	-5.756 (.127)
Duration	-.172 (.076)	.175 (.112)	.967 (.037)
Race[b]			
Black	.428 (.228)	-.151 (.358)	-.133 (.107)
Oriental	.171 (.733)	1.307 (.554)	.279 (.322)
Spanish	-.015 (.399)	-.836 (.744)	.114 (.160)
Female (=1)	-1.008 (.174)	1.505 (.361)	-.777 (.075)
-Loglikelihood[c]	1011.7	727.68	8319.0

Table 7 (Continued)

Panel B: Demographic and Human Capital Effects

Destination State

	Departure for Career Reasons		Departure for Personal Reasons		Promotion Within Company	
Constant	-5.070	(.573)	-7.603	(.779)	-5.878	(.189)
Duration (in grade)	-.006	(.082)	.408	(.130)	1.205	(.040)
Race[b]						
Black	.260	(.238)	-.461	(.382)	-.301	(.106)
Oriental	-.114	(.739)	.926	(.581)	-.175	(.342)
Spanish	.093	(.412)	-1.071	(.760)	.166	(.159)
Female (=1)	-.323	(.191)	1.992	(.382)	-.193	(.081)
Age (in years)	-.006	(.015)	-.037	(.014)	-.035	(.004)
Seniority (months)	-.019	(.002)	-.008	(.002)	-.003	(.001)
Educational Level[d]						
2	.828	(.380)	-.492	(.532)	-.040	(.148)
3	.154	(.336)	-.443	(.409)	.333	(.112)
4 or 5	-.003	(.387)	.146	(.375)	.500	(.125)
6	.768	(.266)	.396	(.320)	.938	(.111)
7, 8 or 9	.400	(.349)	-.100	(.537)	1.094	(.150)
-Loglikelihood[c]	943.8		691.8		8099.6	

Table 7   (Continued)

Panel C: Demographic, Human Capital and Organizational Effects

	Destination State		
	Departure for Career Reasons	Departure for Personal Reasons	Promotion Within Company
Constant	-4.428 (.790)	-7.301 (1.07)	-6.637 (.268)
Duration (in grade)	-.001 (.088)	.430 (.138)	1.246 (.040)
Race[c]			
Black	.319 (.246)	-.519 (.409)	-.375 (.109)
Oriental	-.048 (.811)	.821 (.629)	-.240 (.346)
Spanish	.230 (.437)	-1.205 (.809)	.137 (.161)
Female (=1)	-.412 (.224)	2.145 (.491)	-.105 (.087)
Age (in years)	-.015 (.017)	-.033 (.016)	-.034 (.005)
Seniority (months)	-.018 (.002)	-.008 (.002)	-.003 (.001)
Educational Level[d]			
2	.747 (.408)	-.479 (.586)	.183 (.154)
3	.137 (.349)	-.406 (.464)	.325 (.114)
4 or 5	-.015 (.418)	.148 (.401)	.472 (.127)
6	.732 (.308)	.600 (.366)	.757 (.117)
7, 8 or 9	.429 (.392)	.047 (.586)	.940 (.156)
Job Focus[e]			
1	.471 (.527)	.613 (.881)	-.072 (.203)
2	.310 (.459)	-.660 (.483)	-.831 (.167)
6	.256 (.324)	-.527 (.483)	.084 (.124)
7	-.132 (.392)	.398 (.411)	.050 (.147)
8	.438 (.356)	-.348 (.423)	.278 (.137)
9	.111 (.357)	-.013 (.497)	.259 (.137)
10	-.169 (.373)	.000 (.399)	.062 (.123)
Division[f]			
2	-.477 (.525)	-.921 (.495)	.679 (.180)
3	-.131 (.401)	-1.134 (.403)	.796 (.155)
4	-.372 (.472)	-.642 (.404)	.810 (.160)
5	-.248 (.556)	-.875 (.605)	.528 (.211)
Location (1=Home)	-.443 (.231)	.318 (.290)	-.142 (.089)
-Loglikelihood[c]	937.3	682.8	8051.0

[a] The parameters of the multi-state hazard in equation (1) were estimated by the Method of Maximum likelihood, using the algorithm described in Petersen (1986). Number of employees in grade level 7 is 3235.

[b] Excluded group: White and Native Americans.

[c] The loglikelihood is partioned into three parts, one for each of the three rates. The loglikelihood pieces for the full 15 state hazard are not reported in the Table. They are of no use in performing tests of interests for current purposes.

[d] Excluded group: High School Graduate or less education.

[e] Excluded group: Typist, Figure Clerk, Other Clerk, and Other Codes.

[f] Excluded group: Agency.

Table 8

Estimates of the Parameters of the Multistate Hazard Rate Model
for Destination States 1, 2 and 3 from Salary Grade Levels
12-15[a]
(estimated standard errors in parentheses)

Panel A: Only Demographic Effects

| | Destination State | | |
	Departure for Career Reasons	Departure for Personal Reasons	Promotion Within Company
Constant	-6.486 (.278)	-8.777 (1.05)	-7.066 (.110)
Duration	.021 (.080)	-.064 (.325)	1.191 (.034)
Race[b]			
Black	.814 (.351)	.126 (1.04)	.061 (.160)
Oriental	.666 (.792)	-*[c] (.000)	.266 (.350)
Spanish	.613 (.606)	1.332 (1.08)	-.045 (.295)
Female (=1)	-.362 (.243)	1.595 (.454)	.289 (.075)
-Loglikelihood[d]	1217.3	200.6	12804.1

Table 8   (Continued)

Panel B: Demographic and Human Capital Effects

	Destination State		
	Departure for Career Reasons	Departure for Personal Reasons[e]	Promotion Within Company
Constant	-5.310 (.686)		-5.762 (.193)
Duration (in grade)	.271 (.086)		1.356 (.035)
Race[b]			
Black	-.190 (.375)		-.407 (.160)
Oriental	-.300 (.888)		-.240 (.361)
Spanish	-.027 (.628)		-.418 (.310)
Female (=1)	-.692 (.258)		.226 (.075)
Age (in years)	-.029 (.013)		-.043 (.004)
Seniority (months)	-.016 (.001)		-.002 (.001)
Educational Level[f]			
2 or 3	1.002 (.548)		.178 (.114)
4 or 5	1.011 (.545)		.210 (.137)
6	.533 (.491)		.056 (.097)
7	.900 (.513)		.229 (.122)
8 or 9	.746 (.505)		.500 (.122)
-Loglikelihood[d]	1093.9		12555.0

Table 8   (Continued)

Panel C: Demographic, Human Capital and Organizational Effects

	Destination State		
	Departure for Career Reasons	Departure for Personal Reasons[e]	Promotion Within Company
Constant	-5.438 (.792)		-5.884 (.241)
Duration (in grade)	.286 (.087)		1.380 (.036)
Race[c]			
Black	-.189 (.379)		-.401 (.160)
Oriental	-.241 (.863)		-.122 (.369)
Spanish	-.093 (.633)		-.438 (.315)
Female (=1)	-.686 (.268)		.277 (.077)
Age (in years)	-.034 (.014)		-.050 (.004)
Seniority (months)	-.014 (.001)		-.001 (.000)
Educational Level[f]			
2 or 3	.980 (.553)		.180 (.114)
4 or 5	1.036 (.551)		.199 (.136)
6	.611 (.498)		.026 (.098)
7	.901 (.518)		.181 (.123)
8 or 9	.721 (.514)		.427 (.123)
Division[g]			
2	.342 (.368)		.106 (.126)
3	.362 (.370)		.332 (.122)
4	-1.155 (.519)		.141 (.125)
5	.411 (.367)		.663 (.129)
Location (1=Home)	-.092 (.224)		-.059 (.070)
-Loglikelihood[c]	1082.6		12526.0

[a] The parameters of the multi-state hazard in equation (2) were estimated by the Method of Maximum likelihood, using the algorithm described in Petersen (1986a). Number of employees in grade level 12 is 4838. See Table 1 for further definitions of the variables.

[b] Excluded group: White and Native Americans.

[c] No Orientals made the transition, i.e. left the company from grade levels 12-15 for personal reasons. Hence their rate of departure for personal reasons is 0, and the effect parameter is β(oriental)=-∞ (estimate of minus 28 million). Deleting the term for orientals from the hazard yielded identical results in terms of the other parameter estimates and in terms of the loglikelihood.

[d] The loglikelihood is partioned into three parts, one for each of the three rates. The loglikelihood pieces for the full 15 state hazard model are not reported in the Table, as they are of no relevance for the tests performed in this paper.

[e] Too few transitions were made for this reason to allow estimation of the cause-specific hazard with any reasonable level of stability in the parameter estimates. The estimates varied drastically with small changes in the specification of variables.

[f] Excluded group: High School Graduate or less education.

[g] Excluded group: Agency.

Figure 1

The Cause-Specific Log-logistic Rate as a Function of Duration in a Grade, for two Values of the Effect Parameter of Duration[a]
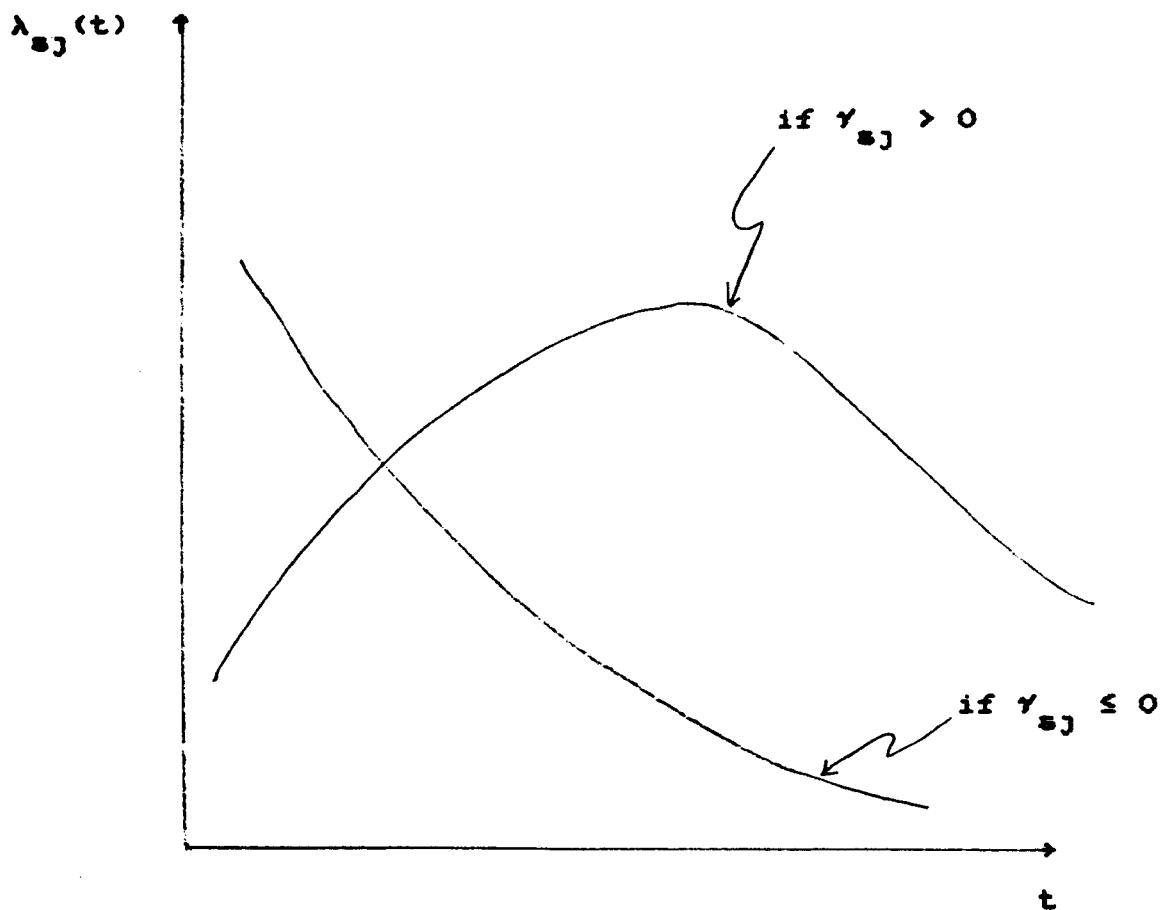
[a] The parameter γ_{sj} gives the effect of duration in state s on the rate for reason j.

Figure 2

The three Cause-Specific Rates in Grade 2 as Functions of
Duration in the Grade[a,b]
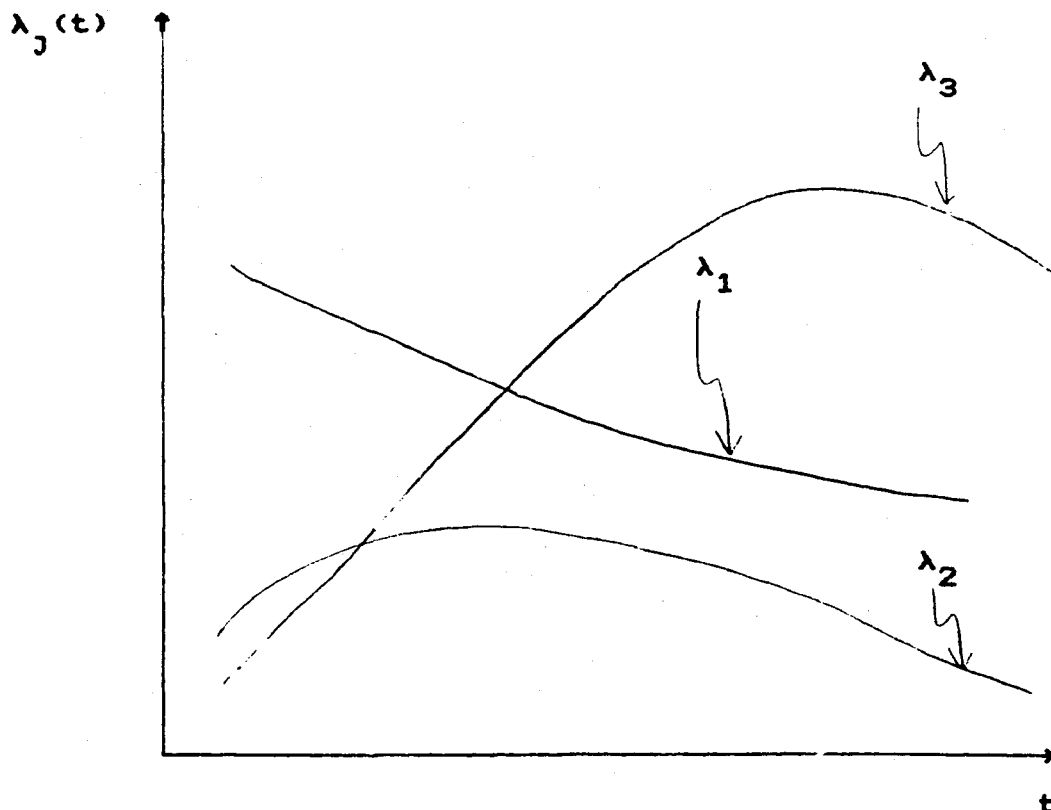
[a] Drawn on the basis of the results in Panel C of Table 5.

[b] The rates λ_1, λ_2 and λ_3 are respectively for leaving the
organization for career reasons, for personal reasons and for
getting promoted from grade 2. See Table 1 for further
definitions.

Figure 2

The three Cause-Specific Rates as Functions of Grade level After One Month Duration in the Grade[a,b]

Panel A: White Males
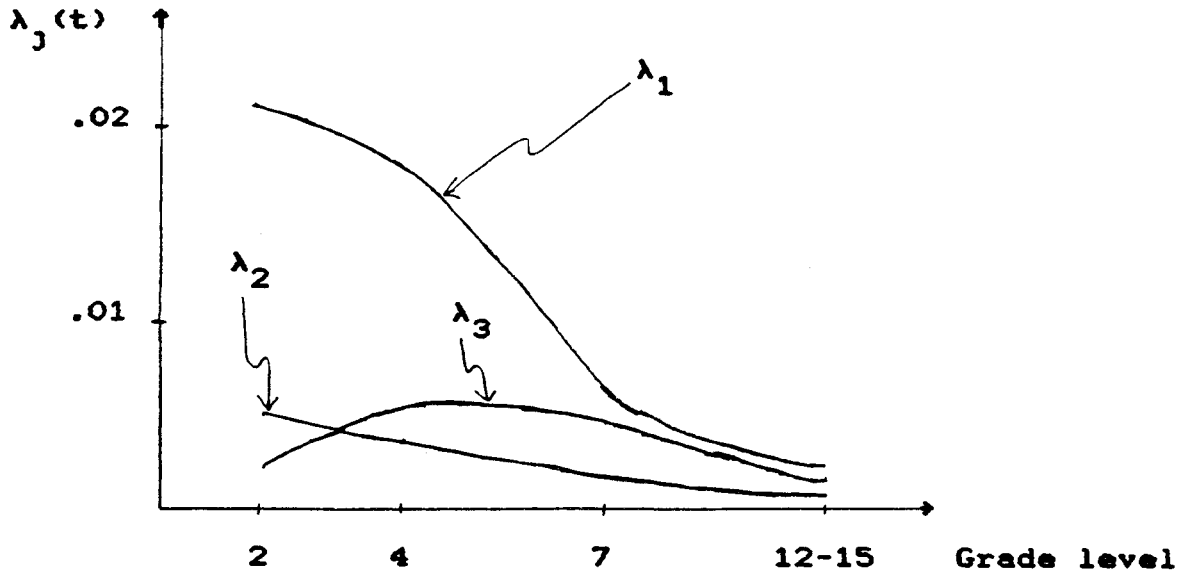
Panel B: White Females:
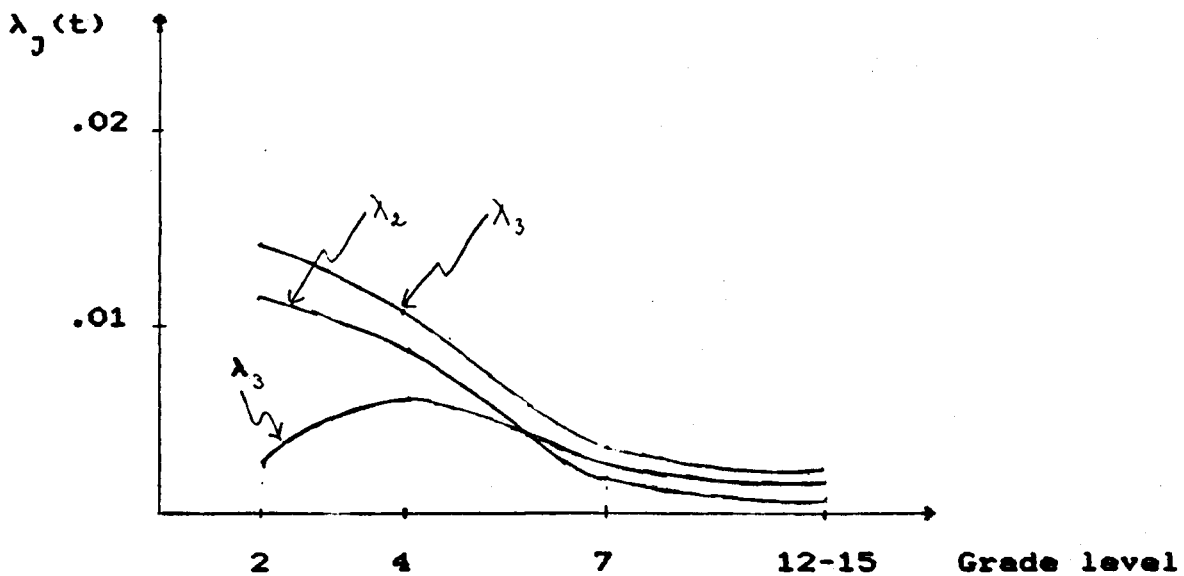
[a] Drawn on the basis of the results in Table 4.

[b] The rates λ_1, λ_2 and λ_3 are respectively for leaving the organization for career reasons, for personal reasons and for getting promoted from the grade. See Table 1 for further definitions.

Departures from an Internal Labor Market[a]

Trond Petersen and Seymour Spilerman[b]

Gender differences in job mobility rates in the United States: A test of human capital and segmentation explanations

David S. Hachen

It is well known that the labor market outcomes for men and women differ significantly. In the United States, women on average earn around sixty percent of what men earn, are concentrated in "female" occupations (Treiman and Hartmann 1981), and are underrepresented in upper level managerial and supervisory positions (Wolf and Fligstein 1978; Wright et al. 1982). However, there are competing explanations of the processes that generate these observed earnings, occupation and class differences. Human capital explanations argue that the behavior of women, in particular the discontinuous nature of their participation in the labor force, accounts for these gender differences (Mincer and Polachek 1974; Polachek 1976, 1979, 1981; Mincer and Ofek 1982). In contrast, segmentation theorists claim that

institutional mechanisms generating gender based occupational segregation are a major source of female disadvantages in labor markets (Bergmann 1974; Wolf and Rosenfeld 1978; Beller 1982; Rosenfeld 1983, 1984). These competing perspectives on gender stratification have been tested in a variety of ways. For the most part past empirical research has either analyzed variation in earnings at one point in time (Mincer and Polachek 1974; England 1982) or earnings growth over time (Mincer and Ofek 1982; Beller 1982; Corcoran, Duncan, and Ponza 1984). However, a few studies, recognizing the importance of mobility processes, have examined gender differences in mobility patterns, most often by analyzing particular transition probabilities using panel data over a limited and short time interval (Wolf and Rosenfeld 1978; Rosenfeld 1980, 1984). It is increasingly recognized, however, that a systematic examination of gender differences in work histories can shed light on the validity of competing explanations of gender differences in labor market outcomes (Treiman 1985).

This research uses job history data and event history analysis to examine gender differences in mobility rates in the United States. In particular this study examines whether the human capital perspective's focus on labor force discontinuities and/or the segmentation perspective's focus on occupational segregation can account for the observed difference between men and women in their rates of upward mobility. In Section I of this paper I present the theoretical arguments of these two perspectives and their hypotheses concerning gender differences in mobility. Section II contains information on the job history data and event history analysis methodologies used to test these hypotheses. In Section III I present the analysis, while Section IV concludes the paper with a discussion of the problems inherent in both human capital and segmentation explanations of gender stratification.

I. THEORETICAL EXPECTATIONS

Both human capital theory and segmentation perspectives have been used to explain gender differences in labor market experiences. The purpose of this section is to draw out the expectations each

perspective has concerning the causes of gender differences in mobility patterns.[1]

Human Capital Explanations: Human capital theory accounts for variation in earnings in terms of labor supply and, in particular, differential investments by individuals in human capital (Mincer 1974; Becker 1975). The basic human capital model has been extended in order to explain gender differences in earnings in terms of the discontinuous nature of female participation in the labor force (Mincer and Polachek 1974). According to this theory, during periods of labor force withdrawal, previously acquired skills depreciate and the returns to one's initial investment decrease. Rational economic actors who anticipate such withdrawals should, therefore, minimize their early investments in depreciable skills and delay such investments until after the occurrence of such discontinuities. Furthermore, when these investments are finally made, the volume of investment should be lower because of the shorter payoff period.

In general, the possibility of depreciating skills should lead individuals, especially women, who expect or experience labor force discontinuities to have less human capital and, therefore, lower earnings and rates of earnings growth. Furthermore, during periods of labor force withdrawal individuals forgo normal appreciation of skills through experience and, therefore, also will have lower amounts of human capital than those with continuous employment patterns.

This human capital explanation of gender differences has been criticized, however, for assuming that the costs of repairing depreciated human capital are the same as the costs of initially acquiring such human capital. Mincer and Ofek (1982) argue that depreciated human capital can be "restored" or "repaired" and, as they demonstrate, the long-run costs of withdrawal are not as great as the short run losses in earnings. Furthermore, Corcoran (1979) and Corcoran and Duncan (1979) argue that this observed wage "rebound" could be the result not

---

[1] For good summaries of different perspectives on gender stratification in labor markets see Rosenfeld (1984) and Corcoran, Duncan and Ponza (1984).

only of the restoration of human capital, but also of improved information on the part of employers and employees which facilitates the correct matching of workers and jobs. As such individuals with labor force discontinuities should experience in the short-run lower wages because of temporary mismatches, but as time in the labor force increases improved information should lead to correct matches and wage increases.

Finally, human capital theory has also been used to explain gender based occupational segregation (Polachek 1981). According to this perspective, occupations differ in their atrophy rates, defined as "the loss of earnings potential when skills are not continuously used." (Polachek 1981, p. 62). Economic actors, and in particular women, who expect discontinuity in the use of their skills should, therefore, choose to enter occupations with lower atrophy rates. Furthermore, occupations with lower atrophy rates are also likely to have lower appreciation rates, i.e. lower returns to experience. Thus Polachek's extension can account for both the lower earnings of those with discontinuous labor force experiences (i.e. women) and the segregation of women in "female" occupations, if it can be shown that "female" occupations have both lower atrophy and appreciation rates.[2]

In general, human capital theory explains gender differences in labor market outcomes in terms of differences among individuals in discontinuous labor force participation rates. Because women are more likely to have such discontinuities, their investments in human capital and their occupational choices should be different from those of men. The end result of these different investment patterns and occupational choices is the observed gender differences in earnings and occupation.

Though human capital theories are explicitly designed to explain variation in earnings, these theories implicitly contain expectations

---

[2] There is a good deal of debate over whether occupations with higher percentages of female workers have both lower atrophy and appreciation rates. Polachek (1981) presents some evidence for this claim, but England (1982) questions his evidence and provides evidence that contradicts Polachek's assumption.

about differences in mobility patterns. If the earnings growth that typically accompanies the appreciation of skills and the acquisition of experience is the result of upward mobility, then labor force discontinuities should be associated with lower rates of upward job moves.[3] Human capital theory specifies a number of mechanisms that could generate this negative association between time out of the labor force and the rate of upward mobility. First, the possibility of a labor force withdrawal should lead individuals to not invest in depreciable skills. Individuals with such lower investments are, therefore, at a competitive disadvantage in pursuing higher-level positions which require such skills. Secondly, the longer one is out of the labor force, the greater the skill depreciation and again the greater the disadvantage in competition for higher-level jobs. Thirdly, forgone appreciation due to labor force withdrawals should result in such individuals being "behind schedule" relative to members of their cohort who have continuous experiences. Such "career" delays should also lower upward mobility rates.[4] Finally, women who expect to withdraw from the labor force will tend to choose occupations with low penalties for withdrawal (i.e. low atrophy rates). However, if occupations with low atrophy rates also have limited prospects for advancement, the occupational segregation resulting from labor force discontinuities can also account for lower female rates of moving up.

In sum, four general hypotheses concerning mobility patterns can be deduced from human capital theory and its extensions. First, both a

---

[3] Of course some growth in earnings does occur without job changing. The argument here is that promotions and advancement are a major source of earnings growth, and therefore, human capital explanations of earnings growth can be used to also explain variation in upward mobility.

[4] In addition employers who invest in their employees are less likely to hire individuals who have been out of the labor force for a substantial period of time because the time period for recouping this investment is shorter for such individuals. If, therefore, the provision of training is related to upward mobility then this phenomenon could explain the lower rates of upward mobility for those with long periods out of the labor force.

withdrawal from the labor force and increases in the length of time of the withdrawal should lower the rate of upward mobility. Secondly, the longer the time since re-entry into the labor force, the higher the rate of upward mobility. Thirdly, the occurrence of a labor force withdrawal, whether short or long in duration, should be associated with higher job mobility and turnover, though as time from re-entry increases job moving should also decrease. Finally, because women are more likely than men to experience labor force discontinuities and to have greater amounts of time out of the labor force, observed gender differences in mobility rates should be substantially reduced when one takes into account differences in the continuity of participation in the labor force.

Segmentation Explanations: Segmentation theories are less analytical and more descriptive than human capital theories, though they have informed a good deal of the research on gender stratification in labor markets. According to this perspective, the segmentation of women into "female" occupations, together with restrictions on mobility across gender-based segments, generates segmented labor markets consisting of non-competing groups. The existence of segmented labor markets not only contradicts a central assumption of neo-classical economics - that the free and unrestricted mobility of labor is a mechanism that brings the supply of labor and the demand for labor into equilibrium - but also can account for gender differences in earnings and earnings growth. In particular, restrictions on female mobility between segments is likely to generate an oversupply of labor for "female" jobs (especially with increases in female labor force participation rates). While in a fully competitive labor market unrestricted mobility would bring supply and demand back into equilibrium, overcrowding in segmented "female" occupations generates significantly lower earnings for women (Bergmann 1974; Blau 1984).

Segmentation theory assumes an asymmetrical restriction on mobility between segments, i.e. female movement into "male" occupations is restricted but not male moves into "female" jobs. There are many views on the nature and causes of such restrictions. Social-

psychological theories concerning internalized stereotypes view such restrictions as the result of women "restricting" their own mobility (Marini and Brinton 1984). Institutionally based explanations tend to focus on overt discrimination (Bergmann 1974; Beller 1982) or statistical discrimination (Thurow 1975; Wolf and Rosenfeld 1978; Baron and Bielby 1986). In the latter case the argument is that employers in "male" occupations are less likely to employ women and invest in them because they expect, on average, lower rates of return on their investments in women, an expectation purportedly grounded in the higher rate of labor force withdrawal for women. Finally, as we have seen, human capital theorists have attempted to explain gender based occupational segregation in terms of the rational choices of women. However, while human capital theory has been used to explain gender differences in initial occupational choice, such theories typically assume occupational continuity and do not attempt to explain why women are restricted from moving between segments.

Whatever the causes of restricted mobility, the asymmetrical pattern of mobility between segments, together with overcrowding and lower earnings has led segmentation theorists to locate "female" occupations in secondary labor markets in contrast to the location of "male" occupations in primary labor markets. According to dual labor market theory, secondary labor markets are characterized not only by lower earnings, extensive competition for jobs, and an oversupply of labor, but also by limited opportunities for upward mobility (Doeringer and Piore 1971). This lower chance of advancement in secondary labor markets is contrasted to primary labor markets characterized by internal labor markets and job ladders where upward mobility is associated with the acquisition of specific skills and experiences (Althauser and Kalleberg 1981). The argument is, therefore, that either the unskilled nature of jobs in secondary labor markets and/or the limited ability to acquire specific skills either through experience or on the job training, underlies the limited nature of upward mobility within secondary labor markets.

Structural explanations of gender differences in labor market outcomes have incorporated the insights of segmentation and labor market theories into their explanations. For example, Wolf and Rosenfeld (1978, p. 827) in an early article on the sex structure of occupations and job mobility, argue that women tend to be excluded from the primary sector and internal labor markets and are, therefore, more likely to occupy positions in what they broadly define as a craft labor market characterized by little on the job training, skills that are not firm specific , and a low likelihood of occupational advancement.

The segmentation perspective on gender differences contains a number of expectations concerning gender differences in mobility patterns. First, this perspective assumes that female mobility into "male" occupations is restricted. This hypothesis has been tested in a variety of ways, and while it has been shown that there is some mobility between segments, there is also evidence that women are less likely to move into "male" occupations (Beller 1982; Rosenfeld 1983, 1984).[5] For our purposes, however, the more important expectation is the hypothesis that gender differences in upward mobility result from the restriction of women to "female" occupations in secondary labor markets with limited opportunities for advancement. According to segmentation theory, individuals occupying jobs in female-dominated occupations should have lower rates of upward mobility. Furthermore, because it is women who are more likely to occupy jobs in "female" occupations, the observed gender difference in upward mobility should be substantially reduced when one takes into account the percent female in an occupation.

While the segmentation perspective has clear expectations regarding upward mobility, its expectations concerning job mobility, turnover and employment stability are less clear-cut. On the one hand, it has been argued that secondary labor markets are characterized by higher levels of job turnover because of the lower

---

[5] Preliminary analysis using the data set discussed below did show that women have lower rates of moving into "male" occupations.

attachment of workers to their jobs and employers (Edwards 1975). On the other hand, it has also been argued that the restrictions on mobility in secondary labor markets lead individuals to become "stuck" in dead-end jobs. It is possible, therefore, that the expected higher turnover in "female" occupations is counterbalanced by the dead-end nature of jobs in these occupations, resulting in the absence of any significant association between percent female in an occupation and the degree of overall job mobility.

In sum, three general expectations emerge from the segmentation perspective. First, the greater the percent female in the occupation of an individual, the lower the rate of upward mobility. Secondly, when one takes into account the percent female in an occupation, the observed gender difference in upward mobility should be substantially reduced. Finally, and more tentatively, percent female in one's occupation should not be related to the rate of overall job mobility.

II. DATA AND METHODS

In order to test human capital and segmentation expectations concerning gender differences in mobility patterns we utilize job history data and event history analysis. The job history data set used in this study contains information on 2788 jobs held by 1498 respondents since 1940 in the United States.[6] These data were collected by Wright (1985a) through a national survey of the U.S. labor force in 1980. While the primary purpose of this survey was to collect detailed information about a respondent's current job in order to estimate the size of various social classes (Wright et al. 1982), less detailed information was also obtained from respondents about their past two jobs. A job change was defined for respondents as a change in employers or a substantial change in

---

[6] Excluded from the analysis are jobs held prior to 1940, military jobs and jobs classified as self-employed. The exclusion of the self-employed is based on the finding in previous research that the determinants of their "job-shifting" are quite different from the determinants of job moves for employees (Hachen 1986a). Of the remaining 2788 job-person matches, 1073 (38.5%) are the "current" jobs of respondents (in 1980) and are treated as censored cases.

their duties and the job title of their job for the same employer.[7] Besides information on the beginning and ending dates for each job, information was also collected on the reasons for leaving a job (quit, laid-off or dismissed, promoted) as well as numerous characteristics of the job such as: whether they were self-employed; whether they were supervisors; whether as supervisors they had any say in the pay, promotions or disciplining of their subordinates; the number of hours worked per week; and standard information on the occupation and industry. In addition for all respondents we know when they entered the labor force, whether after entering the labor force they were ever out of work and not looking for work and, if so, the beginning year of this exit and the length of time out of the labor force.

This data set provides a rich source of information on the work histories of men and women. Not only does it contain event history data on job shifts, but also useful information on labor force withdrawals. As such these data are particularly useful in testing hypotheses concerning gender differences in mobility patterns.

In order to fully utilize this information event history analysis methods are employed (Carroll 1983; Tuma and Hannan 1984; Allison 1984). Because job moves are qualitative changes (events) that occur in time, event history analysis can be used to model the determinants of the timing of such events, i.e. the rate at which various types of job moves occur. The dependent variable in the analysis, therefore, is a hazard rate, defined as the instantaneous probability of a event (e.g. a job shift) occurring in the next movement of time, relative to that period of time:

$$r(t) = \lim_{\Delta t \to 0} \frac{P(t, t + \Delta t)}{\Delta t} \tag{1}$$

In terms of job changes, this hazard rate depicts the instantaneous rate at which people exit jobs. However, in analyzing job

---

[7] Thus within employer job changes are allowed. However, respondents were explicitly told that a promotion in which they did basically the same tasks should not be counted as a job change.

shifts we are not just interested in the rate of job exiting, but in the rate of specific types of job shifts. Type-specific rates can be analyzed with the competing risks model where there are K mutually exclusive and exhaustive types of events (Tuma and Hannan 1984; Allison 1984; Hachen 1986b). These K types of events can be specified in terms of the different reasons for leaving a state (e.g. the distinction between involuntary and voluntary job exits) or different types of transitions (e.g. the distinction between upward, downward and lateral job shifts). In the competing risks model we have, therefore, K type-specific rates defined as:

$$r_k(t) = \lim_{\Delta t \to 0} \frac{P_k(t, t + \Delta t)}{\Delta t} \qquad (2)$$

The value of a type-specific rate depicts the rate at which a particular type of event occurs in a population irrespective of the occurrence of other types of that event (Hachen 1986b).

In this study we restrict our analyses to the rate of voluntary job exits to another job. The other "competing risks", therefore, are involuntary job exits and voluntary exits either out of the labor force or into self-employment. Gender differences in these other types of job exits have been investigated elsewhere (Hachen 1986a). Because our theoretical concern is with the determinants of voluntary movement among jobs, it is logical to so restrict this analysis.[8]

The rate of voluntary moves to another job can be viewed as a general indicator of job stability and is analyzed here in order to test human capital and segmentation expectations concerning overall job mobility. However, we are also interested in analyzing the directionality of job moves and, in particular, the determinants of upward mobility.

Past research has employed a variety of criteria in order to distinguish the direction of job moves. Carroll and Mayer (1986) for example contrast the earnings in destination and origin jobs.

---

[8] The analysis is restricted to voluntary moves to another job not by excluding other types of moves from the data set, but by treating these cases as censored (Allison 1984).

Characterizing upward moves as job shifts were the earnings level is higher in the destination job has a number of advantages. In particular earnings attainment is an important dimension of work histories and is clearly of central concern in both human capital and segmentation theories of gender differences. Unfortunately, this data set does not contain information on the earnings of individuals in their past jobs.[9]

An alternative criterion is change in occupational status. This method of mapping the directionality of job moves is used by Sørensen and Tuma (1981) and has informed research on gender differences in mobility patterns (see Wolf and Rosenfeld 1978; Sewell, Hauser and Wolf 1980; Rosenfeld 1980). However, there is a good deal of debate over the applicability of Duncan's status scores in mapping the position of women in labor markets given that his initial scores were derived from an analysis of male earnings and education (Stevens and Featherman 1981). More importantly, a change in occupational status can only occur when one changes occupations. Upward moves within occupations are, therefore, by definition lateral occupational status moves. This is particularly problematic given the increasing evidence that women and men differ in the probability of upward moves within the same occupation (Baron and Bielby 1985).

Ideally, therefore, one would like a job based measure of status and hierarchical position in the stratification system. One such possible measure is a person's authority position, in particular whether one does or does not occupy an upper-level supervisory or managerial job. Defining upward job moves as job shifts where one has greater authority in the destination job has a number of advantages. First, upward authority moves are likely to correspond to increases in earnings, thus tapping this important aspect of attainment. Secondly there is mounting evidence that gender segregation is not only occupationally based but also job based (Baron and Bielby 1985). As such a job based measure of the directionality of a move is likely to capture more of the

---

[9] One problem however with earnings data is the reliability of retrospective reports on past earnings.

gender difference in upward mobility than an occupationally based measure. Finally, given that authority is one of the central dimensions in recent conceptions of social class (Wright et al. 1982; Wright 1985b), analyzing mobility in terms of changes in authority allows us to address issues of central concern to class analysis theorists.

Operationally, the definition of an upward authority moves is based on the distinction between sanctioning supervisors, task supervisors and workers. Sanctioning supervisors are defined as those with the most authority in that they can both assign tasks and allocate incentives; task supervisors as those with less authority in that they can only assign tasks; and workers as those with no authority in that they do not supervise the work of others.[10] With this tripartite classification, an upward authority job move can be initially defined as any increase in authority.

One problem with such a definition of upward authority mobility is that it artificially imposes a ceiling on who can be upwardly mobile because sanctioning supervisors can not move up. However sanctioning supervisors really are "at risk" to move up in terms of authority. The problem is that our measures of authority do not capture upward mobility among this class of employees. To solve this problem we incorporate two additional pieces of information into our definition: whether the destination job is classified as being in a managerial occupation (1970 occupational census codes 201 to 245) and whether the respondent told us that the reason for the job change was a promotion. Therefore, if a job shift by a sanctioning supervisor is initially classified as lateral, and if the move is either into a managerial occupation or a

---

[10] The distinction between task and sanctioning supervisors is based on the respondent's answer to the question concerning whether as a supervisor they had any say in the pay, promotions or disciplining of their subordinates.

promotion, then the job shift is reclassified as an upward authority move.[11]

In sum, in order to test human capital and segmentation hypotheses, the following analysis takes as the dependent variables two type-specific rates: the overall rate of voluntary job moves to another job and the rate of voluntary upward authority job shifts.

Event history analysis provides a method for specifying and estimating parametric models of the determinants of such rates. Hazard rates can vary either as a function of time in a state (duration dependency) and/or population heterogeneity. Our primary interest is in population heterogeneity, i.e. the effects of differences among people and the jobs they hold on the rates of voluntary and upward authority moves. However, a few words should be said about duration dependency.

Past research on job shift rates has shown that the rate of various types of job moves is a function of time in a job (Sørensen and Tuma 1981). A frequently used functional form for this relation is the Gompertz distribution where the rate of job shifting decreases (or increases) exponentially with time in a job. Our model of job mobility, therefore, includes a parameter reflecting this non-stationary nature of job shift rates.

Returning to population heterogeneity, we model job mobility rates in terms of a set of explanatory variables. Because rates by definition can not be negative, the functional form for this relation is the loglinear specification:

$$\ln r_k(t) = a + b_{k1}x_1 + b_{k2}x_2 \ldots + b_{kn}x_n + c_k(t) \qquad (3)$$

Under this specification the natural logarithm of a particular type-specific rate is viewed as an additive function of the relevant

---

[11] As a result of this modification, 70 of the formerly 701 lateral authority moves by sanctioning supervisors are reclassified as upward authority moves.

explanatory variables and time in the state (job).[12] The parameters (the b's) in this model gauge the effects of the explanatory variables on the natural log of the rate, while the parameter c measures the effect of time in a job on the rate.[13]

The explanatory variables in our models of both the rate of voluntary moves and of upward authority moves can be divided into four sets. Table 1 details how these are defined, while Table 2 contains descriptive statistics for each variable.

[TABLES 1 AND 2 HERE]

The first set consists of four baseline variables which past research has shown are related to job mobility: education, occupational status, time in the labor force, and a dummy variable for whether one worked

---

[12] The loglinear specification implies that the rate is a multiplicative function of the set of explanatory variables, i.e. $r_k(t) = \exp[a]$ $\exp[b_{k1}x_1]$ . . . $\exp[b_{kn}x_n]$ $\exp[c_k(t)]$. The parameters under this specification are the antilogs of the parameters under the loglinear specification, i.e. $\exp[b_{kn}]$.

[13] According to this model the explanatory variables are constants over time. Technically in the following analysis one variable is treated as a time-varying covariate - time in the labor force - in that its value varies not only across individuals, but for each individual over time in a job. The full model used here therefore is of the following form:

$$\ln r_k(t) = a + b_{k1}x_1 + b_{k2}x_2 \ldots + b_{kn}x_n + b_{kn+1}x_{n+1}(t) + c_k(t)$$

part time in the job.[14] Education is measured in terms of academic credentials ranging from 0 (no high school diploma) to 4 (an advance degree). Occupational status, a general indicator of the "goodness" of occupations (Goldthorpe and Hope 1972), is measured using a transformation of Duncan's (1961) original occupational status score developed by Sørensen (1979). This transformation yields a status distribution which approximates an exponential distribution. Time in the labor force is measured as the length of time from entry into the labor force until a job shift, excluding the amount of time, if any, the respondent was out of the labor force.[15] Finally, part-time work is defined as those jobs in which a person worked on average less than 35 hours per week.

---

[14] These variables primarily concern what is called the reward-resource model. Past research (Tuma 1976; Sørensen and Tuma 1981; Hachen 1986a; Carroll and Mayer 1986) shows that individuals with higher resource levels (e.g. education) have higher mobility rates, while those with higher reward levels (e.g. occupational status) have lower rates. In addition time in the labor force can be viewed as an indicator of the discrepancy between individual resources and job rewards in that those in the early parts of their "careers" are likely to hold jobs whose rewards are not commensurate with their resources, while those later in their "careers' are likely to occupy jobs where rewards are more commensurate with resources. As such time in the labor force "picks up" the effects of other indicators of individual resources (e.g. ability) and job rewards (e.g. earnings) that are not included in our models (see Sørensen and Tuma 1981). While we would have preferred to incorporate such other measures into our models (in particular a measure of earnings), this data set does not include such measures. However, as Carroll and Mayer's (1986) research demonstrates, the omission of earnings does not seriously alter the substantive results.

[15] We exclude time out of the labor force from this measure because when one is out of the labor force one is not at risk to experiencing a job move. Furthermore, the measure of time in the labor force is designed to capture the fact that the rate of job moving (and in particular upward mobility) declines over time as one's job rewards become more commensurate with an individuals resources (Sørensen and Tuma 1981). However, during periods of labor force withdrawal this process of convergence of rewards and resources can not occur and, therefore, a more valid measure of time in labor force should exclude the length of time out of the labor force.

The second set of variables pertain to two ascribed characteristics of individuals, gender and race. The variable female is coded 1 for women, while the variable nonwhite is coded 1 for nonwhites (who are disproportionately black). While our primary interest is in gender differences in mobility rates, we include the race contrast in order to control for this possible source of heterogeneity in job shift rates.

The third set of variables pertain to labor force withdrawals. Three variables are included in our models: a dichotomous variable which measures whether a person withdrew from the labor force prior to entering his/her job[16]; a continuous measure of time out of the labor force , and a continuous measure of the length of time from reentry into the labor force until entry into a job.

The final set includes one variable pertaining to segmentation expectations concerning job mobility - percent female in the occupation. The value of this variable ranges from 0.16 percent (industrial engineers) to 98.57 percent (registered nurses).[17]

---

[16] Individuals who experience a labor force withdrawal **after** entry into a particular job are not considered to have been out of the labor force.

[17] Percent female in an occupation is computed using published 1980 census data (U.S. Bureau of the Census 1984, Table 4, pp. 295-304). However, the initial coding of occupational data in the Wright survey employed 1970 census 3-digit occupational codes. Because significant changes in the coding of occupations took place between 1970 and 1980 , it was necessary to map the published 1980 data on percent female (which is reported using the new 1980 codes) into 1970 coded 3-digit occupations. Fortunately the Bureau of the Census (1982) has double-coded a sample of the 1980 census by 1970 and 1980 occupational codes. In this analysis we use, therefore, data on percent female in an occupation in 1980. However, while some of the jobs in this data set were held by respondents in 1980, others were entered prior to 1980. Ideally it would be best to have information on the percent female in an occupation based upon data from a year that is closer to the relevant job shift. Recent work by Rytina and Branchi (1984) shows that while there were some significant changes during the 1970s in the percent female in particular occupations, for most occupations the percent female remained relatively constant. We do not believe that the use of 1980 percent female data has resulted in serious measurement error.

These four sets of variables are included in models of both the rate of voluntary job moves and of upward authority job shifts. As noted in the previous section, both human capital and segmentation theories expect gender differences in mobility patterns and, in particular, in the rates of upward moves. Thus controlling for the baseline variables we expect women to have lower rates of upward authority moves.

However, human capital and segmentation theorists differ in their explanations of such gender differences. According to human capital theory both a withdrawal from the labor force and the length of time of the withdrawal should be negatively related to the rate of upward authority moves, while segmentation theory implies that individuals in occupations with a higher percentage of females will have lower rates of such moves. In addition, extensions of human capital theory (i.e. wage rebound theory) imply that the rate of upward authority moves will increase as the time from reentry into the labor force increases. Furthermore, both human capital and segmentation explanations imply that the initially observed gender difference in the rate of upward authority moves will be substantially reduced when we control for each perspectives relevant explanatory variables.

Finally, both human capital and segmentation theory contain expectations regarding the overall rate of voluntary job mobility. Extensions of human capital theory imply that labor force withdrawals lead to more instability (i.e. higher rates of voluntary moves) because of temporary mismatches, but that as the length of time from reentry increases the rate of job moving should decrease (i.e. a return to greater stability). On the other hand, we have argued that segmentation theory implies that there is no relation between percent female in an occupation and the rate of overall job mobility.

III. ANALYSIS

We proceed to test human capital and segmentation hypotheses concerning voluntary and upward authority job shift rates by estimating parameters in a series of hierarchical models. Model I – the baseline model – stipulates that these rates are a function of the baseline variables (duration in a job, time in the labor force, education,

status, part-time work) and both the gender and race contrasts. Model II - the human capital model - includes three additional variables pertaining to labor force withdrawals: whether one has ever been out of the labor force, the length of time out of the labor force, and the length of time from re-entry into the labor force until entry into the origin job. Finally, in Model III one additional variable - percent female in the occupation - is added in order to test segmentation expectations. The additive parameters for each of these models are estimated under the specification in equation (3).[18] Table 3 contains these parameter estimates for both the log of voluntary and upward authority job shift rates.

[TABLE 3 ABOUT HERE]

Voluntary job shift rates: Turning first to the three models for voluntary moves, we employ a likelihood-ratio chi-square test in order to judge whether the inclusion of additional variables improves the model's fit. Under the null hypothesis of no difference between the two models, this statistic has an asymptotic chi-square distribution with associated degrees of freedom equal to the number of constraints which distinguish the two models. This test statistic is equal to two times the positive difference between the log-likelihoods.

The test for Model I contrasts this model with the null model which assumes a time-constant, population homogeneous rate. For this null model the log-likelihood is -3893.959. The contrast with the log-likelihood for Model I (-3687.446) yields a chi-square of 413 with 7 degrees of freedom (d.f), which is significant at the .05 probability level. The contrast between Models II and I yields a chi-square of 20 with 3 d.f., which is also significant. However, the contrast between Models III and II yields a chi-square of .346 with 1 d.f., which is not significant. Based upon

---

[18] Maximum likelihood techniques are used to estimate these parameters. The procedure used here employs BMDP's non-linear regression routine which allows one to specify a likelihood function. Based on the work of Peterson (1986), a Fortran sub-routine containing the likelihood function is linked to the main program.

these global tests we can conclude that the rate of voluntary moves is a function of the baseline variables, the labor force withdrawal variables, but not the percent female in an occupation.

Examination of the parameter estimates in the top half of Table 3 provides more detailed information. These estimates reflect the additive effects of the explanatory variables on the log of the rate. As we see in Model I, increases in duration in a job, time in the labor force and occupational status result in lower rates of voluntary moves. On the other hand, increases in educational attainment and working part-time (in contrast to full-time) increase this rate. However, neither the gender nor the race contrast is significant. Thus controlling for the baseline variables, men and women do not differ in the overall rate of job shifting.[19]

The magnitude of these effects is ascertained by converting the additive parameters in Table 3 (the b's) into the multiplicative parameters (the antilog of b, e^b). While the value of an additive parameter depicts how much the log of the rate increases or decreases for a unit increase in x, the value of a multiplicative parameter depicts how much the rate is multiplied by for a unit increase in x. Furthermore, the value of

$100(e^b - 1)$ depicts the percentage increase (or decrease) in the rate for a unit increase in x. For example, the value of the additive parameter for part-time work is .5143. The corresponding multiplicative parameter is 1.67, implying that part-time workers' rate of voluntary moves is 67% higher than the rate for full-time workers [$100(e^{.5143} - 1) = .67$].

Turning to the labor force withdrawal variables in Model II we see that the rate of job moving is higher for those who have been out of the labor force (b = .4688), though this rate decreases as the amount of

---

[19] The fact that men and women do not differ in the rate of moving could be due to the inclusion of the part-time variable in Model I. We examined this possibility by excluding the part-time variable (along with education and status) from our baseline model. While the female effect became positive, it was not significant at the .05 probability level.

time out of the labor force increases (b = -.0775). Furthermore, the effect of the length of time from re-entry into the labor force on this rate is not significant.

The human capital expectations regarding job mobility receive, therefore, some qualified support. Exits from the labor force do imply greater job instability in the future, though the degree of instability is decreased by prolonged periods out of the labor force. Furthermore, according to these data this greater instability does not decrease as time from re-entry increases.[20]

Of particular interest is the opposite effects of an exit from the labor force and the length of time of such a withdrawal. For those who have been out of the labor force the rate of voluntary moves is 60% higher than the rate for those with continuous employment patterns $[100(e^{.4668} - 1) = 59.5]$. However, each additional year out of the labor force decreases this rate by about 7.5% $[100(e^{-.0775} - 1) = -7.5]$. It turns out, therefore, that the rate of voluntary moves is the same for those who have been out of the labor force 6 years and those with continuous employment patterns.[21] Furthermore, those with withdrawals greater than 6 years have lower rates than the continuously employed, while those with shorter periods have higher rates.

One possible interpretation of these findings is that while, as expected, labor force discontinuities increase employment instability, those with substantial periods out of the labor force face a different kind of disadvantage in the future. Such lengthy withdrawals may imply the complete loss of previously acquired human capital and, therefore, severe limitations on one's future employment opportunities.

---

[20] Job moving does decrease as time in the labor force increases, and time in the labor force increases as time from re-entry into the labor force increases. However, net of the lower job mobility rate that results from increases in time in the labor force, time from re-entry is not related to this mobility rate.

[21] For those who withdraw from the labor force for 6 years, the total effect of such a withdrawal is .004 [= .4688 + 6(-.0775)], while the total effect for those who have never exited from the labor force is zero.

Individuals with such limited opportunities will tend to be "stuck" in their jobs and their mobility rate will, therefore, be significantly lower than either those with shorter durations out of the labor force or those with continuous employment patterns.

Turning finally to Model III we see, as expected, that the percent female in one's occupation is not related to the rate of voluntary moves. Together with the fact that men and women are similar in this rate, these findings imply that gender is not a determinant of the overall volume of movement.

Upward authority job shift rates: While both human capital and segmentation expectations regarding job mobility are supported to some degree, the situation is quite different for upward authority moves. The bottom half of Table 3 contains information on the determinants of the rate of this type of job shift. According to the log-likelihood statistics, Model I - the baseline model - is a significant improvement over the null model (chi-square of 160 with 7 d.f.). However, Model II, which includes the labor force discontinuity variables, does not improve the fit (the chi-square of 2.8 with 3 d.f. is not significant at the .05 probability level). However, the contrast between Models III and II, which tests for the significance of percent female in an occupation, is significant (chi-square of 5.7 with 1 d.f.). Based upon these global tests we can conclude that this upward mobility rate is a function of the baseline variables and the segmentation variable, but is not related to the human capital variables pertaining to labor force withdrawals.

Turning to the parameter estimates in the bottom half of Table 3 we see in Model I that increases in job duration, time in the labor force and occupational status decrease upward mobility, while more education increases this rate. And while both the part-time and racial contrasts are not significant, net of these other factors women (in contrast to men) do in fact have lower rates of upward authority moves. The value of this negative effect (b = -.3370), implying that the female upward authority mobility rate is about 29% below the male rate $[100(e^{-.3370} - 1) = -28.6]$.

While women do in fact have lower upward mobility rates, the important issue is whether human capital or segmentation variables can account for this observed difference. Turning to Model II we see that none of the labor force discontinuity parameters are significant, and that the female coefficient is virtually the same as in Model I. We conclude that neither exiting from the labor force nor the length of time of such exits affect the rate of upward authority moves. Therefore, human capital explanations do not account for the observed gender difference in this rate.[22]

In contrast, the percent female in one's occupation is related to this rate. However, contrary to segmentation expectations, increases in the percent female in an occupation are associated with higher rates (b = .0050). Furthermore, taking into account the percent female in one's occupation does not reduce the negative net female effect, but increases it (b = -.5379). Thus while women have lower upward mobility rates than men, individuals in "female" occupations (who are more likely to be women) have higher rates than those in "male" occupations.

Inspection of the magnitude of these effects helps to clarify this paradoxical finding. The net effect of being female reduces the rate of upward moves by about 42% $[100(e^{-.5379} - 1) = -41.6]$, while each increase in the percent female in an occupation increases this rate by half of a percent $[100(e^{.0050} - 1) = .5]$. Combining these two pieces of information we find that, net of other factors, a women would have to be in an occupation with over 107% women (a logical impossibility) in order to have the same upward authority mobility rate as a man in an

---

[22] It could be that other "human capital" variables account for some of the observed gender difference in this rate, such as hours worked per week and education. We tested this possibility by contrasting a model which included only duration in a job, female and non-white with our baseline model (model I). The female coefficient in the more restricted model is -.3316 (s.e. = .1044) which is very similar to the coefficient for women in model I (b = -.3370). We conclude, therefore, that women have lower upward authority mobility rates than men and that neither the labor force discontinuity variables nor other "human capital" variables account for this gender gap.

occupation with no women. However, because there are no men in totally "male" occupations, a more informative contrast would compare women in their typical "female" occupation with men in their typical "male occupation.

According to these data, for women the mean percentage female in their occupations is 67%, while for men it is 27%. The total effect, therefore, for a women in the average female-concentrated occupation is -.2029 [= -.5379 + (67*.005)], while the total effect for a man in the average male-concentrated occupation is .135 [= 0 + (27*.005)].[23] Taking the antilog of the difference between these two numbers we find that the rate of upward authority moves for women in their typical "female" occupation is 72% of the rate for men in their typical "male" occupation

$[e^{(-.2029- .1300)} = .72]$. Thus, for the most part, the large negative net female effect outweighs any positive increase in this rate women receive for being in female-dominated occupations.

Not only do the coefficients in Table 3 imply that women on average have lower rates of upward authority moves, but they also imply the following rank order: men in the typical "female" occupation (i.e. 67% female) have the highest rate; men in the typical "male" occupation (i.e. 27% female) the next highest rate; women in the typical "female" occupation the second lowest; and women in the typical "male" occupation the lowest rate.[24] There are two important implications of this rank order. First, within "female" occupations women are less upwardly mobile than men. Secondly, for women who do enter "male"

---

[23] The values of these two total effects are relative to the value of zero for men in occupations with 0% females.

[24] The total effects for each of these groups is: males in the typical female occupation, .3350; male in the typical male occupation, .1350; female in the typical female occupation, -.2029; female in the typical male occupation, -.4029. These values are relative to a zero point which is the total effect for being a male in a zero percent female occupation. These computations assume that there is not a significant interaction between gender and percent female in an occupation. A test for such an interaction effect showed this to be the case.

occupations, their rate of upward mobility is not only lower than men in "male" occupations, but also lower that women who remain in "female" occupations.

In summary, we have seen that neither human capital nor segmentation theory can account for the observed gender differences in upward authority mobility. While women are more likely to withdraw from the labor force and be in "female" occupations, net of these differences women continue to have lower upward authority mobility rates. In the concluding section of this paper I will discuss the implications these findings have for understanding gender stratification in labor markets.

Downward and Lateral Authority Job Shift Rates: While we have no prior expectations regarding the rates of downward and lateral moves, a complete picture of gender differences in job mobility requires a brief discussion of them. The top half of Table 4 contains the parameter estimates for the rate of downward authority moves[25], while the bottom half pertains to lateral moves. In regards to lateral moves, their determinants are virtually the same as those for all voluntary moves.[26] The pattern, however, for downward moves contains some interesting findings.

[TABLE 4 ABOUT HERE]

First, as we see in Model I, among the baseline variables only time in the labor force and part-time have a significant (and negative) effect on this rate. Furthermore, neither the gender nor the race contrast are significant. Secondly, according to Model II exiting from the labor force does increases the rate of downward moves, though increases in

---

[25] Workers are not "at risk" to experience a downward authority move. Therefore, our models of downward mobility are estimated by excluding workers from the analysis.

[26] One difference between the determinants of lateral and all voluntary moves in the female coefficient. For lateral moves this coefficient is positive, while for all voluntary moves it is negative. However, in only one case is this coefficient significant at the .05 level - Model II for lateral moves.

the amount of time out of the labor force decrease this rate.[27] Finally, when we include percent female in an occupation (see Model III) we find that (a) "female" occupations have lower downward mobility rates but that net of this (b) women have higher rates of moving down. It turns out that in this case the rate of downward mobility for women in their typical "female" occupation is 34% higher than the rate for men in their typical "male" occupation.[28]

IV. CONCLUSION

We began this paper by presenting two explanations for gender differences in mobility - the human capital explanation which focuses on gender differences in labor force continuity and the segmentation explanation which focuses on occupational segregation. We have seen that neither of these explanations accounts for the observed gender difference in upward authority mobility, though both labor force discontinuities and occupational location are related to mobility patterns. Labor force withdrawals increase the rates of voluntary, downward and lateral authority moves, though increases in the length of time out of the labor force depresses each of these rates.[29] However, labor force withdrawals are not related to the rate of upward authority moves

---

[27] The implication is that those who have left the labor force for 3.5 years have the same rate of downward moves as those with continuous employment patterns.

[28] The total effect for women in their typical "female" occupation (67% female) is .1244 [= .5398 + (67*-.0062)], while the total effect for men in their typical "male" occupation (27% female) is -.1674 [= 0 + (27* -.0062)]. The antilog of the difference between these two numbers is 1.3388 implying that the downward mobility rate for women in their typical "female" occupation is 34% greater than the rate for men in their typical "male" occupation.

[29] We have also seen that the length of time from re-entry into the labor force has no net effect on any of the mobility rates analyzed in this study.

and, therefore, can not account for the observed gender difference in this rate.[30]

In regards to segmentation theory we have seen that being in a "female" occupation increases the rate of upward authority moves and lowers the rate of downward moves, while the percentage of females in one's occupation is not related to the rates of voluntary and lateral authority moves. Furthermore, women continue to have lower upward authority mobility rates than men even when we control for percent female in one's occupation. Therefore, segmentation theory also can not account for the observed gender difference in upward mobility.

If neither human capital nor segmentation theory can account for the lower female rate of upward authority mobility, then what factors can? In particular, how can one account for the paradoxical finding that women (relative to men) have lower upward authority mobility rates, but those in "female" occupations (relative to "male" occupations) have higher rates?

To begin with, while segmentation theorists may be correct in claiming that women are restricted in their mobility into "male" occupations, they are wrong in assuming that there are limited opportunities for upward mobility in "female" occupations. As we have seen the rate of upward authority mobility is higher in "female" occupations. This may be due to the more extensive nature of supervision in these occupations, which by creating more supervisory and managerial positions increases the opportunity for upward mobility within these occupations. In contrast, work in "male" occupations may be more autonomous and less supervised and, therefore, the opportunity

---

[30] This does not imply that labor force discontinuities can not account for the observed gender differences in career profiles (e.g. age-earnings profiles). A labor force withdrawal could be associated with downward mobility if individuals upon reentering the labor force tend to hold jobs with rewards below those obtained in their jobs prior to exiting from the labor force. Unfortunately, these data do not allow us to test this hypothesis. What our analysis has shown is that contrary to human capital theorists, labor force withdrawals do not affect rates of upward authority mobility after reentry into the labor force.

for advancing into supervisory positions more limited. According to this argument the occupational differences in upward mobility reflect differences in opportunity structures, which in turn are a function of different organizational arrangements for managing and controlling work.[31]

Two caveats to this argument should be made. First, upward movement from female occupations may be into male occupations and not just within female occupations. However, some of this between occupation upward authority mobility is likely to be into male-dominated managerial occupations where a person manages the work of those in a "female" occupation. Secondly, the observed lower upward mobility rate in "male" occupations may be due not only to the less extensive nature of supervision, but also to differences in the types of upward mobility. In "male" occupations upward movement may be from lower-level supervisory positions into higher-level managerial positions, while in "female" occupations such movement may be from non-supervisory positions into lower-level supervisory jobs. Our measure of upward mobility is more likely to pick up the latter type of upward movement and classify the former type as a lateral authority job shift.[32] Thus while different organizational arrangements for managing work may account for the higher upward mobility rate in "female" occupations, further research on the degree and nature of such mobility is needed.

Even however if segmentation theory can be modified to account for the higher upward authority mobility rate in "female" occupations, it can not explain the consistent finding that women have lower rates of

---

[31] This argument is consistent with much of the recent work on labor markets which links organizational arrangements to labor market outcomes. See, for example, Baron and Bielby (1980) and Hodson (1984).
[32] The hypothesis here is that while the opportunity for upward mobility is more extensive in "female" occupations, such mobility is more likely to occur at lower levels in authority hierarchies. In contrast, in "male" occupations, while the opportunity for upward mobility is less extensive, when it does occur it is more likely to be into higher-level managerial positions.

upward authority moves. Segmentation theory argues that observed gender differences in upward authority mobility are accounted for by occupationally-based segregation and that, therefore, men and women in either female or male dominated occupations will have similar rates of upward authority mobility. As such this perspective implies that the path to upward mobility for women is through movement into "male" occupations where the rate of upward mobility is higher.

Our findings not only contradict the claim that upward mobility is higher in "male" occupations, but also the claim that men and women have similar upward mobility rates in both male and female dominated occupations. The finding that women continue to have lower upward authority mobility rates even when we control for percentage female in their occupations, implies that gender segregation is not just occupationally based, but also job and class based.[33] While being in a "female" occupation increases upward authority mobility, women in "female" occupations have lower upward authority mobility rates than men in "female" occupations. Furthermore, when in fact women enter male occupations, they not only have a lower upward authority mobility rate because "male" occupations have lower rates, but they also have a lower rate than men in "male" occupations.

All of this may imply that there exists a dual system of gender stratification in labor markets. Not only are women restricted (for whatever reason) in their movement into "male" occupations, but within whatever occupation they occupy women are also restricted in their movement into positions of authority. Thus even if women were to move into "male" occupations, this need not imply greater gender equality if the class and job based nature of gender stratification persists.

---

[33] Our findings, therefore, are consistent with the growing literature that demonstrates that gender segregation occurs at less aggregated levels than 3-digit census occupational categories. See for example Baron and Bielby's (1985) study where they use job titles instead of aggregate occupational categories in their analysis of gender segregation.

While the existence of a dual system of gender stratification is consistent with the findings of this research, a good deal more research is needed on the complex interrelation between gender, occupation, class and the organization of work. This dynamic analysis of job mobility has shown, however, that both human capital and segmentation explanations of gender differences in labor market outcomes are problematic and, as currently formulated, these theories can not account for the observed gender differences in mobility patterns.

References

Allison, Paul D. 1984. *Event History Analysis: Regression for Longitudinal Data.* New York: Academic Press.

Althauser, Robert P. and Arne L. Kalleberg. 1981. "Firms, Occupations and the Structure of Labor Markets: A Conceptual Analysis." Pp. 119-149 in *Sociological Perspectives on Labor Markets,* edited by Ivar Berg. New York: Academic Press.

Baron, James N., and William T. Bielby. 1980. "Bringing the Firm Back In: Stratification, Segmentation, and the Organization of Work." *American Sociological Review* 45:737-65.

————————. 1985. "Organizational Barriers to Gender Equality: Sex Segregation of Jobs and Opportunities." Pp. 233-51 in *Gender and the Life Course,* edited by Alice S. Rossi. New York: Aldine.

————————. 1986. "Men and Women at Work: Sex Segregation and Statistical Discrimination." *American Journal of Sociology* 91:759-99.

Becker, Gary S. 1975. *Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education.* Second edition. New York: National Bureau of Economic Research.

Beller, Andrea H. 1982. "Occupational Segregation by Sex: Determinants and Changes." *The Journal of Human Resources* 17, 3:371-92.

Bergmann, Barbara. 1974. "Occupational Segregation, Wages and Profits When Employers Discriminate by Race and Sex." *Eastern Economic Journal* 1:103-10.

Blau, Francine D. 1984. "Occupational Segregation and Labor Market Discrimination." Pp. 117-143 in *Sex Segregation in the Workplace,* edited by Barbara Reskin. Washington, D.C.: National Academy Press.

Carroll, Glenn R. 1983. "Dynamic Analysis of Discrete Dependent Variables: A Didactic Essay." *Quality and Quantity* 17: 425-60.

Carroll, Glenn R. and Karl Ulrich Mayer. 1986. "Job-Shift Patterns in the Federal Republic of Germany: The Effects of Social Class, Industrial Sector, and Organizational Size." *American Sociological Review* 51:323-41.

Corcoran, Mary. 1979. "Work Experience, Labor Force Withdrawals, and Women's Earnings: Empirical Results Using the 1976 Panel Study

of Income Dynamics." Pp. 216-245 in *Women in the Labor Market*, edited by Cynthia Lloyd, Emily Andrews and Curtis Gilroy. New York: Columbia University Press.

Corcoran, Mary and Greg J. Duncan. 1979. "Work History, Labor Force Attachment, and Earnings Differences Between the Races and Sexes." *Journal of Human Resources* 14:3-20.

Corcoran, Mary, Greg J. Duncan and Michael Ponza. 1984. "Work Experience, Job Segregation, and Wages." Pp. 171-191 in *Sex Segregation in the Workplace*, edited by Barbara Reskin. Washington, D.C.: National Academy Press.

Doeringer, Peter B. and Michael J. Piore. 1971. *Internal Labor Markets and Manpower Analysis*. Lexington, Mass.: D.C. Heath.

Duncan, Otis D. 1961. "A Socioeconomic Index for all occupations." Pp. 109-138 in *Occupations and Social Status*, edited by Albert J. Reiss. New York: Free Press.

Goldthorpe, Jonathan K and Keith Hope. 1972. "Occupational Grading and Occupational Prestige." Pp. 19-80 in *The Analysis of Social Mobility: Methods and Approaches*, edited by Keith Hope. New York: Oxford University Press.

Hachen, David S. 1986a. "Career Mobility Patterns in the United States: A Dynamic Analysis of Job Shift Rates." Unpublished manuscript. University of Washington.

_____. 1986b. "Interpreting Parameters in the Competing Risks Model." Unpublished manuscript. University of Washington.

Hodson, Randy D. 1984. "Companies, Industries, and the Measure of Economic Segmentation." *American Sociological Review* 49:335-48.

Marini, Margaret M. and Mary C. Brinton. 1984. "Sex Typing In Occupational Socialization." Pp. 192-232 in *Sex Segregation in the Workplace*, edited by Barbara Reskin. Washington, D.C.: National Academy Press.

Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*. New York: Columbia University Press.

Mincer, Jacob and H. Ofek. 1982. "Interrupted Work Careers." *Journal of Human Resources* 17:3-24.

Mincer, Jacob and Solomon W. Polachek. 1974. "Family Investments in Human Capital: Earnings of Women." *Journal of Political Economy* 82:S76-S108.

Polachek, Solomon. 1976. "Occupational Segregation: An Alternative Hypothesis." *Journal of Contemporary Business* 5, 1:1-12.

_____. 1979. Occupational Segregation Among Women: Theory, Evidence, and a Prognosis." Pp. 137-157 in *Women in the Labor Market*, edited by Cynthia Lloyd, Emily Andrews and Curtis Gilroy. New York: Columbia University Press.

_____. 1981. "Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure." *Review of Economics and Statistics* 63:60-69.

Petersen, Trond. 1986. "Estimating Fully Parametric Hazard Rate Models with Time-Dependent Covariates: Use of Maximum Likelihood." *Sociological Methods and Research* 14:219-246.

Rosenfeld, Rachel. 1980. "Race and Sex Differences in Career Dynamics." *American Sociological Review* 45:583-609.

_____. 1983. "Sex Segregation and Sectors." *American Sociological Review* 48:637-55.

_____. 1984. "Job Changing and Occupational Sex Segregation: Sex and Race Comparisons." Pp. 56-86 in *Sex Segregation in the Workplace*, edited by Barbara Reskin. Washington, D.C.: National Academy Press.

Rytina, Nancy F. and Suzanne Branchi. 1984. "Occupational Reclassification and Changes in Distribution by Gender." *Monthly Labor Review* 107, 3:11-17.

Sewell, William H., Robert M. Hauser and Wendy C. Wolf. 1980. "Sex, Schooling, and Occupational Success." *American Journal of Sociology* 86:551-83.

Stevens, Gil and David L. Featherman. 1981. "A Revised Socioeconomic Index of Occupational Status." *Social Science Research* 10:364-95.

Sørensen, Aage B. 1979. "A Model and a Metric for the Intragenerational Status Attainment Process." *American Journal of Sociology* 85:361-84.

Sørensen, Aage B. and Nancy B. Tuma. 1981. "Labor Market Structures and Job Mobility." Pp. 67-94 in *Research in Social Stratification and Mobility.* Vol. 1, edited by Robert V. Robinson. Greenwich, Connecticut: JAI Press.

Thurow, Lester C. 1975. *Generating Inequality: Mechanisms of Distribution in the U.S. Economy.* New York: Basic Books.

Treiman, Donald J. 1985. "The Work Histories of Women and Men: What We Know and What We Need to Find Out." Pp. 213-231 in *Gender and the Life Course,* edited by Alice S. Rossi. New York: Aldine.

Treiman, Donald J. and Heidi I. Hartmann, eds. 1981. *Women, Work, and Wages: Equal Pay for Jobs of Equal Value.* Washington D.C.: National Academy Press.

Tuma, Nancy B. 1976. "Rewards, Resources and the Rate of Mobility: A Non-Stationary Multivariate Stochastic Model." *American Sociological Review* 41:338-60.

Tuma, Nancy B. and Michael T. Hannan. 1984. *Social Dynamics: Models and Methods.* Orlando: Academic Press.

U.S. Bureau of the Census. 1982. *1970-1980 Census Comparability: Detailed Occupation Sorted by 1970 Codes.* Unpublished Preliminary Data. Washington D.C.: U.S. Department of Commerce.

U.S. Bureau of the Census. 1984. *1980 Census of Population, Volume 2, Subject Reports Occupation by Industry.* Washington D.C.: U.S. Department of Commerce.

Wolf, Wendy C. and Neil D. Fligstein. 1979. "Sex and Authority in the Workplace: The Causes of Sexual Inequality." *American Sociological Review* 84:232-52.

Wolf, Wendy C. and Rachel Rosenfeld. 1978. "Sex Structure of Occupations and Job Mobility." *Social Forces* 56:823-44.

Wright, Erik O. 1985a. *The Comparative Study of Class Structure and Class Consciousness: Public Use Codebook.* Ann Arbor, Michigan: Inter-University Consortium for Political and Social Research.

_____. 1985b. *Classes.* London: New Left Books.

Wright, Erik O., C. Costello, D. Hachen and J. Sprague. 1982. "The American Class Structure." *American Sociological Review* 47:709-26.

Table 1.   Definitions of Variables

Variable	Definition
Time in the labor force	Calculated from date of entry into the labor force until entry into the job, less time out of labor force, if any.[a]
Education	Number of academic credentials: 0 = no h.s. diploma; 1 = h.s. diploma; 2 = some college; 3 = college degree; 4 = advance degree.
Status	Sørensen (1979) transformation of Duncan's occupational status score.
Part-time (hours)	1 = worked on average less than 35 hours per week in job; 0 = worked 35 hours or more per week.
Female	1 = female; 0 = male.
Nonwhite	1 = nonwhite; 0 = white.
Out of labor force	1 = withdrew from labor force prior to job; 0 = no such withdrawal.
Time out of labor force	Length of time out of labor force.[b]
Time from re-entry	Calculated from date of re-entry into labor force until date of entry into job.[b]
% Female in occupation	Percent female in 3-digit census occupation (range: .16% -industrial engineers to 98.57% - registered nurses).

[a] In analyses time in the labor force is treated as a time-varying covariate.

[b] Coded 0 for those with no such withdrawal.

Table 2: Descriptive Statistics (n = 2788)

Variables	All	Women	Men
Time in the labor force			
mean	9.17	8.76	9.50
(st. dev.)	(9.20)	(9.06)	(9.30)
Education			
mean	1.99	1.92	2.05
(st. dev.)	(1.08)	(1.01)	(1.13)
Status			
mean	1.12	1.08	1.15
(st. dev.)	(.99)	(.79)	(1.12)
Part-Time			
percent	14.6	21.4	9.1
Female			
percent	44.6	--	--
Nonwhite			
percent	14.0	14.8	13.3
Out of labor force			
percent	16.0	24.0	9.6
Time out of labor force			
mean	.49 [3.08] [a]	1.03 [4.31]	.06 [.63]
(st. dev.)	(2.20) [4.71]	(3.18) [5.31]	(.37) [.10]
Time from re-entry			
mean	.69 [4.28]	1.06 [4.43]	.385 [3.99]
(st. dev.)	(2.61) [5.21]	(3.20) [5.29]	(1.96) [5.07]
% Female in occupation			
mean	44.65	67.02	26.66
(st. dev.)	(31.86)	(27.23)	(22.56)

[a] Numbers in brackets are means and standard deviations for those who have exited from the labor force prior to entry into job.

Table 3. Estimates of Models for Rates of Voluntary and
Upward Authority Job Shifts[a]

Variables	Model		
	I	II	III
A. Voluntary Moves (baseline rate r - .0912)			
Duration	-.0719*** (.0084)	-.0713*** (.0085)	-.0715*** (.0085)
Time in the labor force	-.0254*** (.0037)	-.0232*** (.0038)	-.0241*** (.0038)
Education	.1634*** (.0318)	.1552*** (.0321)	.1519*** (.0324)
Status	-.1703*** (.0378)	-.1613*** (.0379)	-.1598*** (.0380)
Part-time (hours)	.5142*** (.0781)	.5228*** (.0783)	.5186*** (.0785)
Female	-.0565 (.0602)	-.0314 (.0614)	-.0614 (.0781)
Nonwhite	-.0667 (.0879)	-.0795 (.0882)	-.0790 (.0882)
Out of labor force	---	.4688*** (.1146)	.4698*** (.1146)
Time out of labor force	---	-.0775*** (.0203)	-.0757*** (.0204)
Time from re-entry	---	-.0227 (.0186)	-.0226 (.0186)
% Female in occupation	---	---	.0007 (.0012)
Constant	-1.8832	-1.9187	-1.9338
Log-likelihood	-3687.446	-3677.272	-3677.099
B. Upward Authority Moves (baseline rate r - .0330)			
Duration	-.0809*** (.0150)	-.0815*** (.0151)	-.0826*** (.0151)
Time in the labor force	-.0142* (.0062)	-.0127* (.0063)	-.0117 (.0063)
Education	.3146*** (.0552)	.3119*** (.0554)	.2877*** (.0558)
Status	-.1471*** (.0570)	-.1444* (.0573)	-.1340* (.0577)
Part-time (hours)	.2123 (.1551)	.2177 (.1557)	.1879 (.1567)
Female	-.3370** (.1053)	-.3349** (.1085)	-.5379*** (.1352)
Nonwhite	-.2450 (.1615)	-.2434 (.1618)	-.2384 (.1619)
Out of labor force	---	.3457 (.2013)	.3509 (.2017)
Time out of labor force	---	-.0220 (.0317)	-.0230 (.0320)
Time from re-entry	---	-.0442 (.0320)	-.0441 (.0322)
% Female in occupation	---	---	.0050* (.0021)
Constant	-3.1830	-3.2094	-3.3186
Log-likelihood	-1751.789	-1750.391	-1747.536

[a] Standard errors in parentheses. Rates are measured in years.
N - 2788 job-person matches. For definitions of variables see Table 1.

* .01 < p ≤ .05.

** .001 < p ≤ .01.

*** p ≤ .001.

Table 4. Estimates of Models for Rates of Downward and
Lateral Authority Job Shifts[a]

Variables	Model I		Model II		Model III	

A. Downward Authority Moves (baseline rate r - .0305)

Variables	I		II		III	
Duration	-.0190	(.0170)	-.0172	(.0170)	-.0146	(.0172)
Time in the labor force	-.0351***	(.0099)	-.0346***	(.0100)	-.0360***	(.0102)
Education	-.0466	(.0792)	-.0699	(.0807)	-.0476	(.0823)
Status	-.1160	(.0882)	-.0846	(.0887)	-.1111	(.0895)
Part-time (hours)	.8208***	(.2183)	.8112***	(.2164)	.8440***	(.2165)
Female	.2222	(.1555)	.2979	(.1569)	.5398**	(.1872)
Nonwhite	.3087	(.2053)	.2617	(.2067)	.3070	(.2107)
Out of labor force	---		.7030*	(.3434)	.7047*	(.3432)
Time out of labor force	---		-.2357**	(.0895)	-.2342**	(.0897)
Time from re-entry	---		-.0067	(.0711)	-.0109	(.0716)
% Female in occupation	---		---		-.0062**	(.0031)
Constant	-2.8191		-2.8670		-2.7379	
Log-likelihood	-752.948		-747.964		-746.007	

B. Lateral Authority Moves (baseline rate r - .0444)

Variables	I		II		III	
Duration	-.0919***	(.0129)	-.0898***	(.0130)	-.0900***	(.0130)
Time in the labor force	-.0333***	(.0055)	-.0332***	(.0057)	-.0331***	(.0057)
Education	.1146*	(.0465)	.1040*	(.0468)	.1007*	(.0470)
Status	-.3022***	(.0566)	-.2902***	(.0568)	-.2892***	(.0568)
Part-time (hours)	.6924***	(.1058)	.6963***	(.1060)	.6918***	(.1062)
Female	.1484	(.0863)	.1723*	(.0878)	.1392	(.1132)
Nonwhite	-.0570	(.1247)	-.0746	(.1249)	-.0743	(.1248)
Out of labor force	---		.4509**	(.1552)	.4526**	(.1552)
Time out of labor force	---		-.0847**	(.0297)	-.0850**	(.0298)
Time from re-entry	---		-.0020	(.0248)	-.0019	(.0249)
% Female in occupation	---		---		.0008	(.0017)
Constant	-2.3490		-2.3800		-2.3959	
Log-likelihood	-2134.444		-2128.098		-2127.998	

[a] Standard errors in parentheses. Rates are measured in years.
For lateral moves n - 2788 job-person matches. For downward moves n - 1044.
For definitions of variables see Table 1.

* .01 < p ≤ .05.

** .001 < p ≤ .01.

*** p ≤ .001.

Effects of marriage and childbirth on women's labor force participation. A dynamic analysis of immediate events and ensuing states

Angelika Tölke

I. INTRODUCTION

Women's labor force participation as well as other events in the life course are mainly determined by two dimensions:

- first, the historical context in both past and present provides a general setting for the life course. The historical context includes for example social norms and values, legal regulations, the labor market and the state of the economy in general.
- and second, living conditions and events people are experiencing as individuals in the past and in the presence also shape the life course. Past experiences include family origin, educational achievement as well as personal crises; the presence concerns the position in the family life cycle and characteristics of a current job.

Women's (and men's) lives are influenced by these two dimensions at the same time as they take part in producing them.

This paper stresses the individual level in the presence. I'm going to pick up a very small part of women's lives and will limit the analysis of labor force participation to marital and fertility related events and states. I will analyse how and to what extent current family related events and situations are influencing women's labor force participation, in particular the process of leaving employment.

While comparing three different birth-cohorts I capture the extent of change in a certain historical period, too.

We know from previous research that there is a relationship between women's labor force participation and their marital and fertility status. Being married and having small children reduces the labor force participation to a considerable amount. In comparison to that group, single or divorced women are employed to a higher extent whereby cohort effects must be taken into consideration (see Schwarz 1981, Sorensen 1983, Handl 1984)[1] .
In earlier analyses no attempts were made to distinguish between the effects of events and states of the family life cycle -that is the difference between the event of a childbirth and the state of being a parent- are not separately taken into account. This is largely due to the lack of appropriate data. Research questions and the interpretation of the results concerning labor force participation and family life cycle sometimes have been formulated in relationship to events (Kohler/Reyher 1970,Stegmann 1976)[2] .

In this analysis the main research questions are as follows:
In which direction and to which extent is young women's attachment to the labor force altered:

- by entering marriage and by entering parenthood? Effects of these events can be classified as immediate effects on labor force participation
- by the state of beeing married and the state of having at least one child? These state related effects can be classified as ensuing effects
- and are there any indications for a change in the effects of these events and states on the process of leaving employment in the period after the second world war?

---

1. We also know that for example the characteristics of the husband's job or his attitude towards women's labor force participation have an influence on women's employment behavior.

2. Beyond this the causal ordering between fertility and labor force participation is complex and there are investigations with differing results and conclusions. (Sweet 1973, Cain 1966, Stolzenberg/Waite 1977)
    Recent studies based on longitudinal data enabled us to distinguish between events and states and their respective effects of employment behavior. With those data sets we get a closer view on the relationship of labor force participation and marital and fertility history (Cramer 1980, Felmlee 1985).
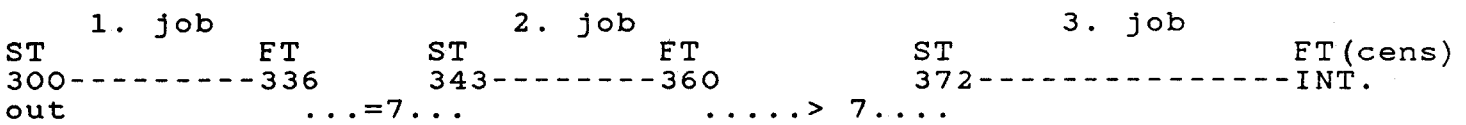
II. DATA

My analysis is based on retrospective life history data, which were collected in the project 'Lebensverlaeufe und Wohlfahrtsentwicklung' between 1981-83[3] . Interviews were completed with 2171 adults from a random sample of the German population belonging to one of three birth cohorts (born 1929-31, 1939-41 or 1949-51). This research utilizes the female subsample of 1086 cases. Fully comparable information on life cycle events for all three cohorts is available up to age 30 (the minimum age of the youngest cohort at the time the data were collected).

The unit of analysis is shifted from individual cases to employment spells. An employment spell is defined as a time period in the labor force with no interruption exceeding seven months[4] .
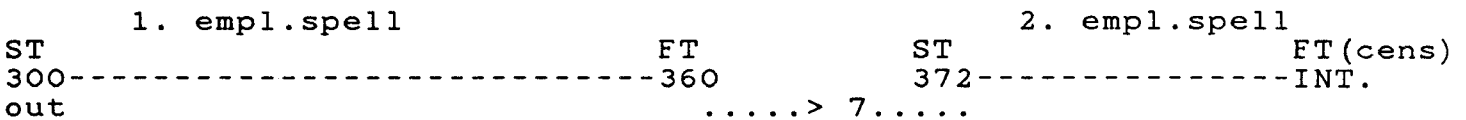The transformation from seperated job spells to employment spells is shown in the following figure 1:

Figure 1

1. Job history before transformation (time in months since 1900)

```
        1. job              2. job                   3. job
ST              FT      ST              FT      ST                    FT(cens)
300--------336         343--------360         372----------------INT.
out            ...=7...               .....> 7....
```

2. Employment spells after transformation

```
        1. empl.spell                                2. empl.spell
ST                                          FT      ST              FT(cens)
300--------------------------------360             372--------------INT.
out                         .....> 7.....
```

ST = starting time   FT = finishing time   INT = time of interview

III. STATISTICAL MODEL

Given space limitation I need to assume the reader's familiarity with the basic terminology of stochastic processes used in the analysis of event data[5] .

The analysis employs the proportional hazard model (Cox 1972, 1975). It assumes that the rate of occurrences changes over time without specifying any particular parametric form. It serves to analyse the effect of covariates on the (time dependent) rate.

The instantaneous rate of the occurrence of an event -the transition from being employed to

---

3. This project is part of the Sonderforschungsbereich 3 which is financed by the Deutsche Forschungsgemeinschaft.

4. THe decision not to treat interruptions up to seven months as employment exits is caused by legal regulations concerning maternity leave. Before 1979 the maternity leave covered three months after birth. Since 1979 women who give birth to a child are entitled to leave their job after a birth for a period of six months without any disadvantages besides less income in that period; they are still formally employed and after that period they are entitled to the same job. Women of the youngest cohort thus had a chance to use this new regulation. In addition to this survival analysis indicated that the difference between leaves of three, six or twelve months are very small.

5. As a general reference see Tuma/Hannan 1984 or in German Andress 1985, Diekmann/Mitter 1984).

being not employed- at a given point of time in a certain employment spell is:

$$r(t) = q(t) \exp[a_1 x_1 + a_2 x_2 + \ldots]$$

or

$$\ln r(t) = \ln q(t) + a_1 x_1 + a_2 x_2 + \ldots$$

where t is duration of employment, $q(t)$ is an unspecified duration-dependent "nuisance function", the x's are the covariates, and the a's are the effects of these variables.

This model assumes that the covariates do not change over time; that means that their values remain the same for the whole employment period, they are fixed. This is a common assumption; for example if we were interested in the effects of background variables or effects of educational level, then we know that these variables are fixed in the past -as characteristics of parents or living conditions in childhood or youth- or that only very few people experience a change for example in the educational level after leaving school (at least in West-Germany).

But we are interested in estimating the effects of the occurance of marital and fertility events and changes in its states the point in time must be included when there is a change in the covariates. That means we are expecting a change in the risk of leaving employment as a result of marital and fertility events during a certain spell of employment.

Including time dependent covariates in the proportional hazard model the equation changes :

$$r(t/x(t)) = q(t) \exp[a_1 x_1(t) + a_2 x_2(t) + \ldots]$$

$$= q(t) * (\exp a_1)^{x_1(t)} * (\exp a_2)^{x_2(t)} * \ldots$$

where the x(t)'s are now time dependent covariates. Using the multiplicative form it is easily seen that the antilogs (exp ai) of the parameters represent the multiplicative factor by which the rate changes due to a unit change in the corresponding covariate (ceteris paribus).

$$\exp a_i^{(x_i+1)} = \exp a_i^{(x_i)} * \exp a_i$$

In our model the nuisance function is assumed to be the same for all members in a particular population. Even when the proportional hazards assumption is violated, it is often a satisfactory approximation (Allison 1984).

IV. STATES AND EVENTS

As we know from theoretical approaches and empirical research in sociology as well as in economics (cf. role theory and human capital theory) traditional definitions of the parental role and traditional division of labor within a family assign women primary responsibility for housework and childcare. Entry into marriage and even more entry into parenthood pushes women in the direction of assuming traditional sex roles. We expect that the effect is decreasing over time but nevertheless still present among the youngest cohort.

Both the immediate event and the longer lasting state are supposed to cause changes in women's labor force participation and even in men's life[6].

Marriage and childbirth come as events well anticipated. Clearly pregnancy proceeds birth, giving at least several months to make plans accordingly. Marriage may be more spontaneous and changes may be less drastic (e.g. in cases were the couple lives together in a premarital arrangement); in general however marriage and childbirth bring about a time of transition.
The occurrance of events means a change in everyday life where new routines have to be developed and learned.

In our model we will relate any occuring change to the event (of marriage or childbirth) if it occurs in a certain period around the event. The event indicators are constructed as follows:

- the period of four months before and four months after a wedding (defining a dummy variable 'MARRIAGE')
- the period of six months before -that means the woman is in the third months of her pregnancy- and the period of four months after a childbirth (defining a dummy variable 'BIRTH').

Thereafter we consider the transition period to be terminated and the woman to be in the state of being married or being a parent (what is indicated by two other dummy variables). Figure 2 illustrates the construction of the various time dependent dummy variables (more elaborate example see appendix A).
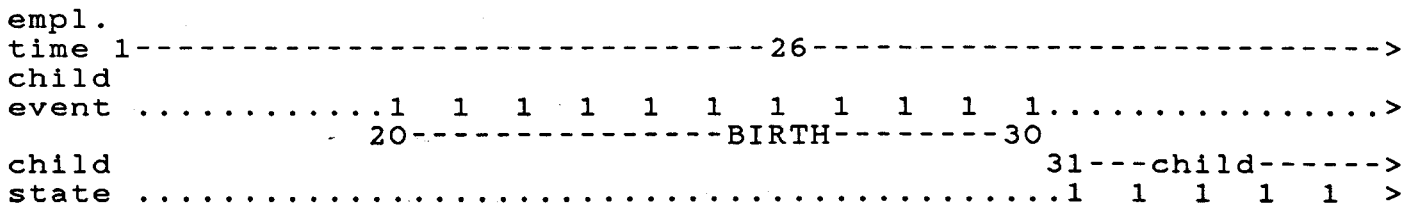
Figure 2

Example 1 (simplified)

Entering marriage and being married

```
empl.
time 1----------------------8---------------------------------------->
marriage
event ........1  1  1  1  1  1  1  1  1...........................>
               4-------MARRIAGE-------12
marriage                                      13-----married---------->
state ........................................1 1 1 1 1 1 1 1 1 1 1 1 >
```
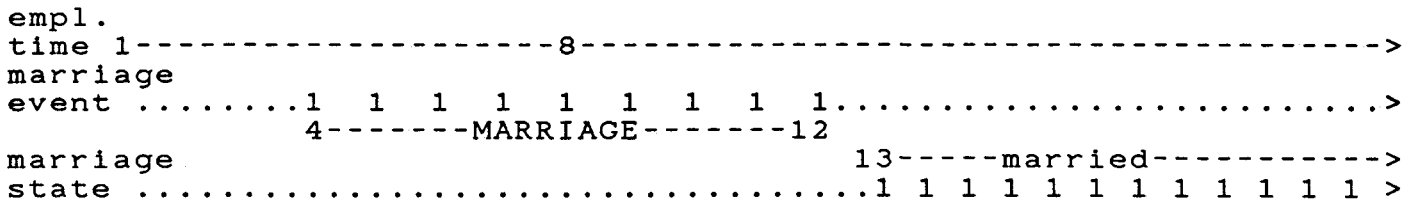
Example 2 (simplified)

Entering parenthood and having a child (being single)

```
empl.
time 1--------------------------26-------------------------->
child
event ...........1  1  1  1  1  1  1  1  1  1  1.................>
              -  20---------------BIRTH--------30
child                                         31---child------>
state ........................................1  1  1  1  1  >
```

The dashed time line corresponds to the time of employment; it starts with the first month of each employment spell and runs to the end of the employment spell -which means that a break of more than seven months takes place, or to the age of 30. In the last case the employment spell is considered censored.

The dots of the dotted lines symbolize zero values that means there is no event, or the woman is

---

6. However men are not assumed to leave the labor force; rather there is an increase in the amount of working hours (see Haggstrom et al 1984) and it has also been suggested that traditional family life has a positive effect on men's long term careers.

not in the state being married (having a child).

As is well known the event and state of marriage and having a child are not stochastically independent. Given mutual dependence over time we expect their effects on employment not to simply add up rather we anticipate interacting effects. These interacting effects can be modeled in a number of ways, details are described in the excursus below.

Excursus: Design Matrix

We consider two trichotomous variables relating to marriage and childbirth with the categories defined as:

```
= 1 state
= 2 event
= 3 none
```

With each variable two dummies are associated:

```
marr = 1 being married (state)      child = 1 having a child (state)
     = 0 else                             = 0 else

   M = 1 wedding (event)               B = 1 childbirth (event)
     = 0 else                             = 0 else
```

As indicated above a simple additive model does not aedequately reflect the data. Including interaction terms in a straightforward way creates some estimation problems. Subpopulations #4 and #7 (see table 1 below) refer to unwed mothers, low in absolute numbers. To circumvent the estimation problems subpopulations #4 and 5 as well as #7 and 8 were combined, i.e. dummy variables representing interaction effects were constructed that do not differentiate between those subpopulations. In particular:

```
Bchild = (B + child) * (1 - marr)
```

To test for differences in structure between the three cohorts, they were jointly estimated with additional dummy variables representing the cohorts:

```
c3 = 1   for cohort 1929-31      c5 = 1   for cohort 1949-51
   = 0   else                       = 0   else
```

Also, interaction terms between cohorts and the marriage/child related dummy variables were included. The partial design matrix (relating to each of the three cohorts) is described in the table below:

Table 1: Partial Design Matrix

Marriage	child	const	marr	M	marr*child	marr*B	M*Bchild	Bchild
1. state	state	1	1	0	1	0	0	0
2. state	event	1	1	0	0	1	0	0
3. state	none	1	1	0	0	0	0	0
4. event X	state	1	0	1	0	0	1	1
5. event X	event	1	0	1	0	0	1	1
6. event	none	1	0	1	0	0	0	0
7. none X	state	1	0	0	0	0	0	1
8. none X	event	1	0	0	0	0	0	1
9. none	none	1	0	0	0	0	0	0

It may help to briefly characterize the columns in the design matrix in substantive terms:

```
M           = marriage
M*Bchild    = Marriage and Birth,or Marriage and having a child
marr        = being married, having no child
marr*child  = being married and having a child
marr*B      = being married, having a birth
Bchild      = childbirth or having a child, being single
```

Finally, it needs to be noted that additional covariates were included to guard against 'spurious effects' of the marriage and childbirth indicators we presently focus our attention upon. These covariates are: respondent's and husband's occupational level and mother's labor force experience .

End of Excursus

V. RESULTS OF THE PROPORTIONAL HAZARD MODEL - EACH COHORT SEPARATELY

a. Methodological remarks concerning the interpretation of the coefficients

In figure 3 the exponential coefficients for each cohort are shown . They give an impression of the effects of marital and fertility events and states and their relative importance for each cohort. Subsequently we will have a look at the changes between the cohorts.

The results are graphically represented in figure 3 below. This tree-like diagram displays the different (multiplicative) factors determining the rate (see section on the statistical model). Since the single factors are estimated simultaneously, the order of presentation reflects the researcher's choice based on substantive considerations, not a stepwise estimation process. Some of the values attached to the arrows can be found directly in table 2 (see appendix B), others are products of these coefficients. For example the value of 14.18 reflecting the contribution of a wedding (cohort 1929-31) results from 22.15 (the overall factor of wedding - dummy variable M) times 0.64 (the differential factor of wedding in this cohort - dummy variable c3M).

In general, the effect shown is the product of all coefficients for which the corresponding dummy variable (column in the design matrix) will be 1. Since the middle cohort was used as reference group all but one effect in the middle part of figure 3 can be found directly in the table of coefficients. The one exception is the effect of having/getting a child when a wedding is imminent. Here two coefficients have to be multiplied given the particular form of the design matrix (the coefficients for Bchild and M*Bchild).

b. Employment exits for each cohort

As we expected variables concerning marriage and parenthood [8] have a tremendous effect on the process of leaving employment (controlling for
respondent's and husband's occupational level, mothers labor force experience). In each cohort events are increasing the risk of leaving employment to a much higher degree than states do.

- Women born 1929-31

The rate of leaving employment for women of the oldest cohort who experience the event of marriage is 14.18 times as high as for single women of that cohort who don't have that event. The risk of leaving employment also increases to an immense amount when there is an immediate responsebility for childcare at the time of the wedding (the woman already has a child or she is pregnant). The risk for this group is 19.11 times as high as for women who don't have any of these immediate obligations when they marry.

In comparison women who are in the state 'being married', that means at least five months after a wedding, have employment exit rates which are 2.12 times as high as their single counterparts.

The risk of leaving employment increases again when married women give birth to a child; it is 16.66 times higher than for those who are married but don't have a childbirth. Married women who have a child which is more than four months old are to a considerable degree at a lower risk to quit than those women who just experienced a childbirth. Their risk is 2.25 times as high as for their married counterparts who have no childcare responsibilities.

---

8. Only first marriage and childbirth were considered.

The risk of leaving employment for unwed women who give birth to a child or who are a parent is about the same as for those single women who don't have any child.

- Women born 1939-41

We know from previous analysis the cohort born 1939-41 had a strong 'family orientation' (in a demographic sense) caused by period effects in the 1960th (Papastefanou 1986, Toelke 1986a, Toelke 1986); women of that cohort are in comparison to the older and younger cohort on the average younger when they give birth to their first child and have more children at the age of 30. This 'family orientation' we recognize in our models too. The risk of employment exits is in all familiy concerned situations higher than in the older and younger cohort -except the state 'being married' and 'being single and having a child'.

The effects of events are again generally higher than the effects of states. For example the risk for employment exits caused by the event 'MARRIAGE' is 22.15 times as high as for single women, the risk for those who are in the state 'married' is just about 2.5 times as high as for their single counterparts.

The first childbirth -while being married- increases the rate for an employment exit about 49 times, whereas the state -being already a parent for some time- increases it just about 7 times. Even the child-effect turns to be significant for single women in this cohort.

- Women born 1949-51

In the youngest cohort -women born 1949-51- we again learn that events are more important for employment exits than states.

And although we don't know yet the significance of the change between the cohorts we see decreasing effects of the event marriage and we recognize increasing child related effects compared to the oldest cohort.
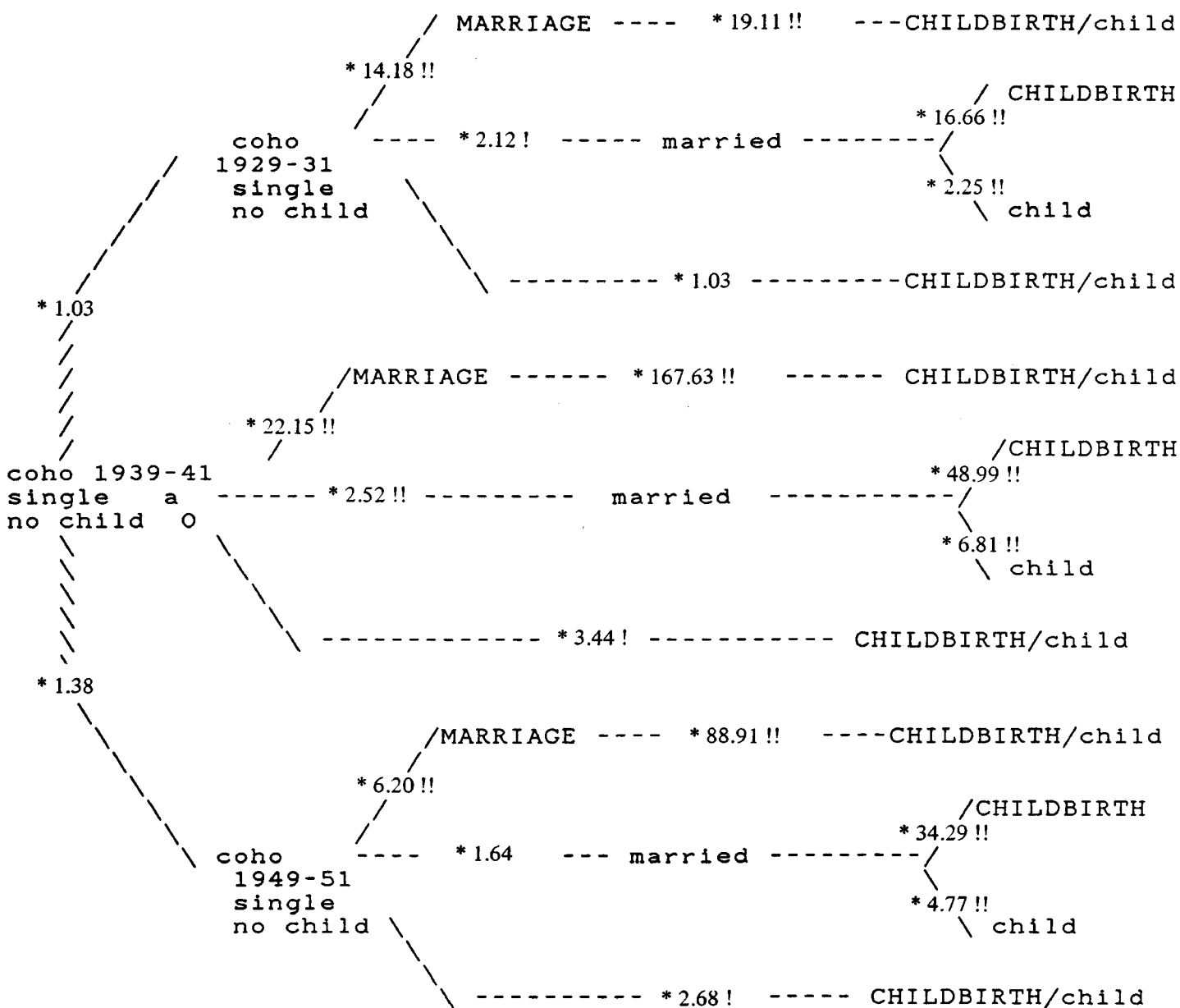
Women who are just going to marry leave employment at a rate that is 6.2 times as high as women who are single. The rate of leaving employment being already married for some time is 1.64 times higher than being single, but it is no longer significant.

If a married woman of that cohort gives birth to a child it increases the chance for employment exits about 35 times; in comparison the risk for married mothers is 4.77 times as high as for those who are married without any childcare responsibilities.

Unwed women who give birth to a child or have a child have a risk of 2.68 to leave employment and it is significant. We suppose that slightly increasing numbers of unmarried couples and improved welfare regulations may have caused it.

Figure 3

Marriage and fertility effects on the process of leaving employment for different birthcohorts

```
                            / MARRIAGE ---- * 19.11 !!    ---CHILDBIRTH/child
                           /
                     * 14.18 !!
                         /                                   / CHILDBIRTH
              /         /                                   * 16.66 !!
             /     coho      ---- * 2.12 !  ----- married -------/
            /      1929-31                                   \
           /       single       \                             * 2.25 !!
          /        no child       \                            \ child
         /                          \
        /                             \ --------- * 1.03 ---------CHILDBIRTH/child
     * 1.03
       /
      /                    /MARRIAGE ------ * 167.63 !!  ------ CHILDBIRTH/child
     /                    /
    /              * 22.15 !!
   /                    /                                        /CHILDBIRTH
  /                    /                                      * 48.99 !!
coho 1939-41          /
single       a ------ * 2.52 !! --------- married -----------/
no child     o  \                                           \
                 \                                           * 6.81 !!
                  \                                           \ child
                   \
                    \ ------------- * 3.44 ! ----------- CHILDBIRTH/child
   * 1.38
      \
       \                   /MARRIAGE ---- * 88.91 !!   ----CHILDBIRTH/child
        \                 /
         \           * 6.20 !!
          \              /                                   /CHILDBIRTH
           \            /                                 * 34.29 !!
            \  coho    ---- * 1.64  --- married --------/
               1949-51                                  \
               single                                    * 4.77 !!
               no child \                                 \ child
                         \
                          \ ---------- * 2.68 ! ----- CHILDBIRTH/child

significant at the level of
   .05  => !    .01  => !!
```

VI. RESULTS OF THE PROPORTIONAL HAZARD MODEL - CHANGE BETWEEN COHORTS

a. Methodological remarks

To test for differences between the cohorts we need to look at the interaction terms between the cohort dummies and the marriage/childbirth indicators. Figure 4 displays these coefficients. Given the construction of the cohort dummies the middle cohort is omitted; the coefficients reflect the difference to this middle cohort. With the exception noted above (child effect in case of a wedding) all coefficient are taken directly from the table of coefficients (see Table 2, appendix B).

b. Extent of change since the second world war

Let us first take a look at changes in the employment behavior of single women. The rate of leaving employment for unwed women who have no responsibilities for childcare is not increasing between birthcohort 1929-31 and 1939-41 and not to cohort 1949-51. This doesn't support a first general guess that the rates of leaving employment will decrease.

If we look at marital and fertility effects there are just four significant changes in the period after the second world war.

The event marriage (without any immediate childcare obligations) is loosing its importance for the process of leaving employment between cohort 1939-41 and 1949-51. This effect is highly significant and we expected it because of increasing numbers of premarital couples -for whom the event 'MARRIAGE' doesn't bring any big immediate change- and because of changes in expectations to married women.

Beyond this norms are no longer as distinctively as they were in the past.

There are three more significant changes which took place in the 1950th and 1960th (pertain to women of cohort 1929-31 and 1939-41).

The effect of a childbirth, being a parent (being married) and being pregnant while marrying increased significantly. This means that women of the oldest cohort who remained employed when they married had a significant lower rate of employment exits when they gave birth to a child and even when they already have been a mother for some time than women of cohort 1939-41. We suppose a higher economical necessity for women of the oldest cohort for a participation in the labor force.

On the other hand women of cohort 1940 experienced the so called 'Wirtschaftswunder' in the 60th. This 'Wirtschaftswunder' gave families a better economical starting point and basis. Thus these women could decide with less economical pressure -norms still present- what they wanted to do.

VII. CONCLUSION

We have investigated the relationship between family related events and states on the process of leaving employment. Our findings point to the importance of differentiating between events and states of marriage as well as parenthood. It seems that women of younger cohorts are delaying employment exits in the family life cycle.
The effect of marriage is decreasing whereas childcare related reasons are of increasing importance.
But although we see some significant changes between the cohorts it is hard to specify a clear trend. Results and interpretations of employment behavior of women must be embedded into the historical context.

Figure 4

Extent of change between cohorts

```
                              / MARRIAGE ---- * 0.11 !!   ----CHILDBIRTH/child
                             /
                      * 0.64
                           /                                    /CHILDBIRTH
                         /                                  * 0.34 !!
             coho     ----- * 0.84 ---- married ---------/
             1929-31                                        \
      /      single   \                                  * 0.33 !!
    /        no child   \                                      \ child
   /                      \
  /                         \ ---------- * 0.30  ------- CHILDBIRTH/child
 * 1.03
 /
/
coho 1939-41
single    a
no child  O
 \
  \
 * 1.38
    \
     \                        / MARRIAGE ------ * 0.53 ------ CHILDBIRTH/child
      \                      /
       \              * 0.28 !!
        \                  /                                   /CHILDBIRTH
         \               /                                 * 0.70
          \    coho    ------ * 0.65 --- married ---------/
             1949-51                                        \
             single    \                                  * 0.70
             no child    \                                     \child
                           \
                             \ ---------- * 0.78 ------ CHILDBIRTH/child
```
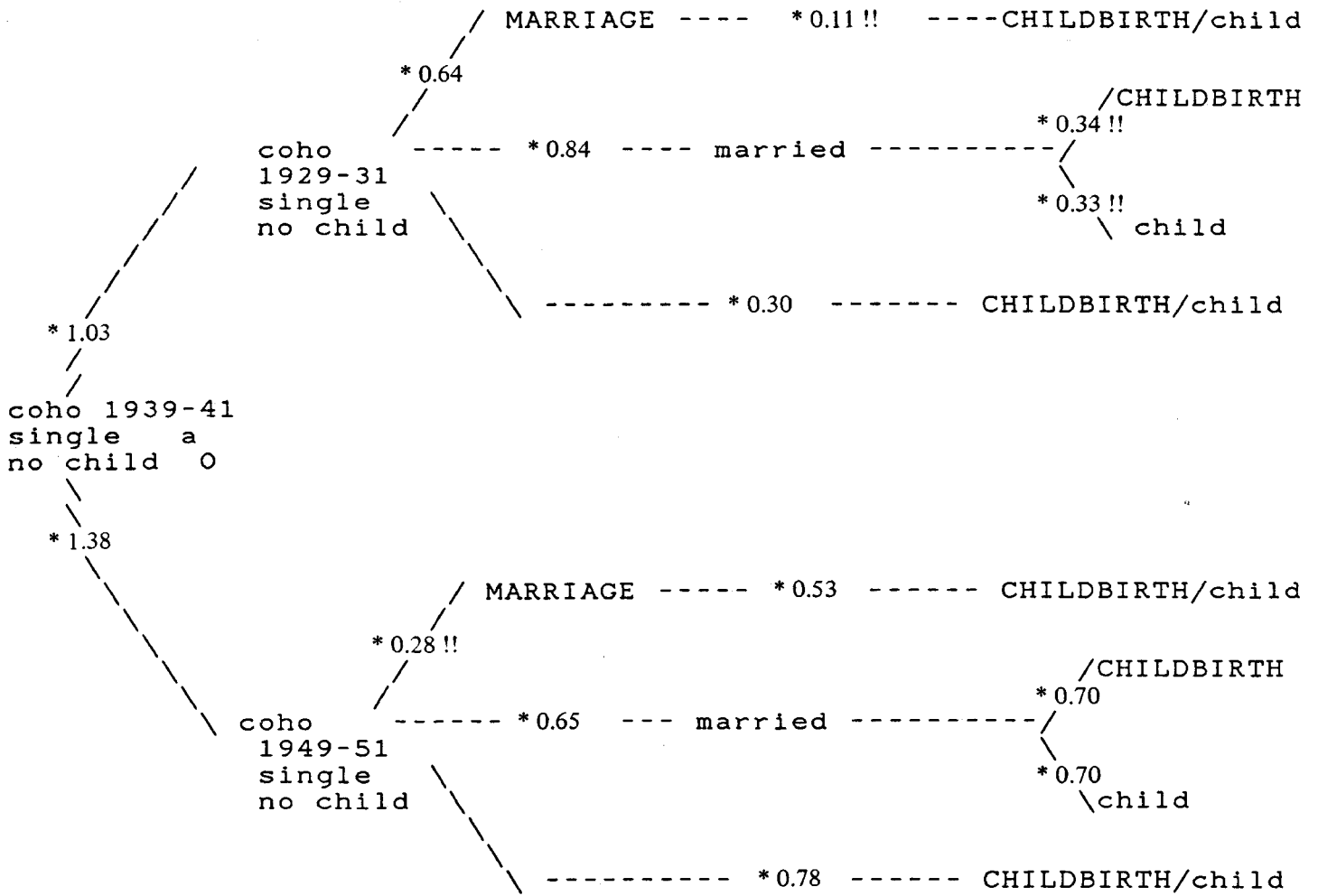
significant at the level of
.05   =>   !
.01   =>   !!

Appendix A

Example 3

Having no child and not being pregnant while entering marriage

```
time 1----------10----------------------------40---------------->
M .........111111111...........................................>
          6-------14
marr...................111111111111111111111....................>
                      15---------------33
Bmarr.........................................1111111111..........>
                                             34-------44
marrchild.........................................111111111111>
                                                 45----------->
MBch .......................................................>.
Bchild......................................................>
```


Appendix B

Table 2
Partial-likelihood-estimates of the effects of marital and fertility
variables the rate of leaving employment

Variable	sign*	Coefficient	Exp(coeff.)
c3		0.03	1.03
c5		0.32	1.38
marr	!!	0.93	2.52
marrchild	!!	1.92	6.81
M	!!	3.10	22.15
MBchild	!!	3.89	48.73
marrB	!!	3.89	48.99
Bchild	!	1.24	3.44
c3marr		−0.18	0.84
c3marrchild	!!	−1.10	0.33
c3M		−0.45	0.64
c3MBchild	!!	−0.96	0.38
c3marrB	!!	−1.09	0.34
c3Bchild		−1.22	0.30
c5marr		−0.43	0.65
c5marrchild		−0.35	0.70
c5M	!!	−1.28	0.28
c5MBchild		−0.38	0.68
c5marrB		−0.36	0.70
c5Bchild		−0.24	0.78

* significant at the level of
  .05 => !   .01 => !!

References

Allison, P., 1984: Event History Analysis. Regression for Longitudinal Event Data. Beverly Hills: Sage Publication

Andress, H.-J., 1985: Multivariate Analysen von Verlaufsdaten. ZUMA-Methodentext, Bd.1. Mannheim

Cain, G., 1966: Married Women in the Labor Force: An Economic Analysis. Chicago: University of Chicago Press

Cox, D. R., 1972: Regression models and life tables, in: Journal of the Royal Statistical Society (Series B), Vol. 34, S.187-202

Cox, D. R., 1975: Practical Likelihood, in: Biometrika 62, 2, S.269-276

Cramer, J. C., 1980: Fertility and female employment: problems of causal direction, in: American Sociological Review 45:167-90

Diekmann, A., Mitter, P., 1984: Methoden zur Analyse von Zeitverlaeufen. Stuttgart: Teubner Studienskripten

Felmlee, D. II., 1985: The Interdependence of Women's Fertility and Labor Force Participation, paper presented at the 80th Annual Meeting of the American Association, Washington D.C.

Haggstrom, G. W. et al, 1984: CHanges in the Lifestyles of New Parents. Santa Monica: Rand Corporation

Handl, J., 1978: Ausmass und Determinanten der Erwerbsbeteiligung von Frauen, in: Beitraege zur Arbeitsmarkt- und Berufsforschung, Nr. 31

Kohler, H./Reyher, L., 1970:Erwerbstaetigkeitsphasen der Frauen, in: MittAB 3, S. 286-297.

Papastefanou, G., 1986: Veraenderungen der Familienbildung in der Bundesrepublik Deutschland seit dem 2. Weltkrieg, unveroeffentlichtes Manuskript, Berlin

Schwarz, K., 1981: Erwerbstaetigkeit der Frau und Kinderzahl, in: Zeitschrift fuer Bevoelkerungswissenschaft, Jg. 7, 1, S. 59-86

Sorensen, A. M., 1983: Women's Employment Pattern after Marriage, in: Journal of Marriage and the Family, vol. 45, 2, p. 311-321

Stegmann, 1976: Unterbrechung und Wiederaufnahme der Erwerbstaetigkeit von Frauen, in: WiSta, 4, S.236-239

Stolzenberg, R./Waite, L., 1977: Age, fertility expectations, and plans for employment, in: American Sociological Review 48, S.769-782

Sweet, J. A., 1973: Women in the Labor Force. New York: Seminar Press

Toelke, A., 1986a: Zentrale Lebensereignisse von Frauen. Veraenderungen im Lebensverlaufsmuster in den letzten 30 Jahren, in: Brose, H. G. (Hg.) Berufsbiographien im Wandel, Opladen

Toelke, A., 1986: Historische Ausgangssituation und Veraenderungen im Ausbildungs- und Erwerbsverhalten junger Frauen in der Nachkriegszeit, in: Voges, W. (Hg.), Methoden der Biographie- und Lebenslaufforschung, 1986, im Druck

Tuma, N. B./Hannan, M. T., 1984: Social Dynamics: Models and Methods, San Diego: Academic Press

Employment sector and unemployment processes

Aage B. Sørensen

INTRODUCTION

Unemployment is an important and carefully monitored social
indicator. The level of unemployment in a society is
generally considered a measure of the social and economic
welfare of a society and it is, therefore, of considerable
political significance. By this indicator, most Western
societies have not been doing well in recent years. Unem-
ployment rates have remained stubbornly high, not only in
the US where unemployment traditionally is high, but also in
many European countries where unemployment was low to very
low in the sixties and early seventies.

Traditionally, unemployment is considered a variable in
macro-economic theory to be manipulated through measures
that affect aggregate demand. The main measure of unemploy-
ment, the so-called unemployment rate, is a statistic that
measures how many people are affected by unemployment and
this measure may well be the relevant variable for the
macro-economic theory. However, it is not a measure that
informs much about individual unemployment processes. The
measure is a proportion formed by counting in the numerator
the number of people looking for work at a point in time
(the survey week) and in the denominator the total number of
people in the labor force[1]. The measure is not a rate in
the technical sense of the term unless it can be assumed
that all unemployment durations are of equal length.

The number of people looking for work in a particular week
is a count of the number of unemployment spells in existence
in that week. The spells are created by people leaving or
being forced out of their jobs. They are terminated by
people reentering employment or leaving the labor force.
The count of spells or people unemployed at a point in time

will be determined by the rate at which spells are created
and by the rate at which they are ended. If the creation
and termination of unemployment spells are completely
determined by the demand for labor and if the labor market
is homogeneous, then aggregate demand should indeed govern
the number of spells to be observed at a point in time and
the usual method of measuring unemployment is adequate.

If, on the other hand, the creation and duration of spells
are governed by other forces than the business cycle, then
the unemployment rate as conventionally measured is not
informative about the mechanisms that govern unemployment;
nor is aggregate demand policies likely to be effective in
reducing employment. If unemployment processes are deter-
mined by individual characteristics, or if the labor market
is not homogeneous, the aggregate demand policies risk to
create excessive demand for some groups or in some sectors
and result in wage inflation, while failing to reduce the
unemployment for other groups or in other sectors of the
labor market.

The experiences of the last decade suggest that the
conception of unemployment processes as homogeneous may not
be adequate because the macro-economic policies that go with
this conception seem to have failed to reduce unemployment.
In fact, the concern for inflation seems by now to have led
to an almost complete abandon of any attempt to directly
reduce unemployment in the aggregate in many countries. The
wage inflation is what one would expect to result from
aggregate demand increases in a heterogeneous labor market
with a great deal of variation in individual unemployment
processes. In fact, the labor market surveys used to
measure unemployment, such as the CPS, show a great deal of
individual and labor market variation in the proportion
unemployed.

Theory that would explain individual variation in unemploy-
ment comes mainly from micro-economics. Particularly
prominent is a view of unemployment as being generated by an
individual's search for better jobs in a highly dynamic
employment system. In search models, persons choose to be
unemployed in order to engage in productive search for
better jobs. Unemployment is a part of the natural turnover
in the labor market (e.g. Hall, 1970). Unemployment compen-
sation and welfare benefits make it economically possible
for individuals to quit their jobs rather than having to
search for better jobs in their spare time. The view
implies that unemployment is a transient and voluntary
phenomenon not effectively dealt with by increasing aggre-
gate demand. It is especially likely to be engaged in by
young people experimenting with their employment and by
those groups where the economic losses of engaging in search
of the job are relatively small: those who may have alter-
native sources of support, for example married women, and
low paid workers who are not much worse off on welfare or
unemployment compensation.

The conception of unemployment as search is consistent with
some of the main features of the distribution of the
unemployed according to demographic characteristics and also
consistent with an increase in overall levels of unemploy-
ment as income support programs become more generous. The
conception is not consistent with the fact that a high
proportion of those who report themselves as unemployed also
report that they did not voluntarily quit their jobs, but
were laid off. Some of those may be argued to not be unem-
ployed at all: the temporarily laid off who have jobs they
will return to (Feldstein, 1975). The rest of the laid-off
can still be said to be voluntarily unemployed arguing from
contract theory that they choose to sign employment con-
tracts that ensure them high and fixed wages in partial
compensation for uncertain employment (Azariadis, 1975).

Both search theory and contract theory implies that
unemployment spells should be of short duration. There is
considerable evidence for a preponderance of short unemploy-
ment spells from research using longitudinal data. In fact,
spell lengths will be overestimated from cross-sectional
surveys of the unemployed reporting on their completed
lengths of spells such as the CPS (Salant, 1977). Still,
the theory and the spell distribution that is consistent
with the theory may understate the welfare consequences of
unemployment as argued by Clark and Summers (1979) who show
that most of the experience of unemployment is felt by a few
who have very long spells of unemployment. This result
dampens the "benign" view of unemployment as a short term
experience widely shared and in fact not completely unavoid-
able in a dynamic economy.

A more comprehensive view of the unemployment process than
offered by search and contract theory should be able to
account not only for short spells, but for the overall
distribution of spell lengths. This means a focus on the
dynamics of the unemployment experience. It seems difficult
to provide such a more comprehensive view without taking
into account that people may change during their unemploy-
ment spells in a manner that increases or decreases their
chances of reemployment. This suggests investigations of
the manner in which length spells depend on variables that
change over the spell as reflected in the form of duration
dependence of the unemployment process. Such an investiga-
tion is the purpose of this paper.

The micro-economic theory sees unemployment as a voluntary
individualistic affair in sharp contrast to the undifferen-
tiated and involuntary view of unemployment assumed in
macro-economic theory. The micro-economic theory pays very
little attention to the specification of the labor market
context for unemployment processes that would condition the

individual level processes. It may be surmised that processes generated by search and processes resulting from anticipated layoffs are unlikely to occur in the same labor markets. Still, the emphasis in the economic theory is exclusively on individual choice rather than on the social structures that condition these choices. Similarly, if indeed the overall length distribution of spells reflects individual change, the sources of such change cannot be identified in the micro-economic theory. These sources presumably have something to do with the labor market context for the unemployment processes, especially the opportunities offered by different labor market structures.

Sociologists of the labor market have in recent years done a great deal of research on labor market structures using a variety of conceptions: dual economy sectors, internal versus external labor markets, primary versus secondary labor markets. This research has rarely focussed on unemployment, with the exception of Schervish (1983)[2]. Instead the research has continued the traditional emphasis of sociological research on socio-economic attainment processes. However, a proper specification of labor markets structures for the study of career and attainment processes should also have implications for the study of unemployment processes. Such a conceptualization and a specification of its relevance for the study of unemployment, in particular the dynamics of the process of ending unemployment, will be presented in the next section.

UNEMPLOYMENT DURATIONS AND LABOR MARKET STRUCTURES

Much sociological research on the labor market is concerned
with identifying labor market variables that cause variation
in socio-economic attainment. These efforts produce a great
deal of descriptive knowledge about the relative importance
of individual versus "structural" characteristics for
observed level of attainments, where these "structural"
characteristics are a variety of industry, firm, and
organizational variables. While this research is useful for
the identification of the places where people are likely to
receive a higher than a lower pay, the research has been
less useful in increasing our understanding of how labor
market structures interact with individual attributes in
producing attainments (Sørensen, 1983). This understanding
is more likely to be obtained from the study of the
mechanisms that create labor market processes in different
labor market structures. It will be a study of mechanisms
for the linking of people and their characteristics to jobs
or positions providing different levels of earnings and
other rewards.

A distinction that appears to be especially useful for the
identification of different mechanisms of labor market
processes is a distinction between closed and open employ-
ment relations. Originally proposed by Weber (1968), the
distinction is elaborated and applied to employment rela-
tions in Sørensen (1977a, 1983). Open employment relations
are here defined as employment relations that are freely
available to anyone with the needed qualifications. In
particular, the access to open employment jobs is not
constrained by the employment decisions of those in jobs at
a particular moment of time as open employment relations
will be of short duration for specific tasks. These are the
employment relations assumed in the neo-classical theory

that sees labor markets as competitive and basically similar
to markets for other goods. In such markets, competition
establishes prices that are wage rates and tied to the
productivity of individuals in the manner described by
marginal productivity theory. For given demand schedules,
equally productive persons will obtain equal wage rates.
Only if there are barriers to mobility creating unequal
demand schedules or lack of perfect information will unequal
wage rates for identical people be observed.

Closed employment relations are available to new employees
only if the present incumbents of jobs have decided to leave
their positions, either for a better job or for retirement.
Employment relations tend to be of longer and indefinite
durations because they can only be reestablished when
vacancies occur in jobs. Closed position systems emerge for
reasons of technology and/or contractual problems in the
manner suggested by internal labor market theory (Doeringer
and Piore, 1971, Williamson, 1975). In internal labor
markets wage rates and other rewards are characteristics of
jobs and not of individuals as administrative arrangements
and not market competition establishes the match between
individual characteristics and job rewards. Changes in wage
rates reflect changes in jobs and are not necessarily tied
to changes in individual productivity as is the case in
competitive systems. Historically and organizationally,
specific mobility regimes established by promotion systems
establish the correspondence between individual
characteristics and jobs, so that identical individuals
usually will not receive identical wage rates even if there
was no uncertainty.

The vacancy competition mechanism created in closed
employment relations and the wage competition mechanism
created in open systems produce very different labor market
processes. Still, it is possible to show that both

mechanisms can be used to account for the main features of observed attainment processes (Sørensen, 1979, 1984). Thus both mechanisms would generate the typical shape of the experience earnings profile observed. However, in wage competition, growth in the early years reflects increases in productivity brought about by on-the-job training and experience, while in vacancy competition, growth reflects the utilization of mobility opportunities established by promotion schemes without necessarily reflecting growth in productivity.

Closed and open employment systems also generate very different unemployment processes. Ideal type versions of both systems should not generate any unemployment at all. In completely open systems where there is perfect information and no state intervention everyone should be able to find employment at some wage rate immediately. In completely closed systems nobody should be forced to leave their jobs. When unemployment occurs in open systems, it is a result either of the existence of minimum wage laws that prevent some from being employed at the wage rate that corresponds to their productivity or because there is less than perfect information. The latter situation is the one assumed in micro-economic search theory. Because information is costly to obtain in terms of time and other resources, people choose to leave their jobs in order to devote their time to search for better jobs. The situation is one where demand chocks in product markets to the labor market equilibria established by competition produce opportunities for finding better jobs. People set an aspiration level and continue to search until they receive a wage offer that matches their aspiration level or reservation wage. The search period corresponds to a period of unemployment. The decision to engage in search should be heavily dependent on the individual's reservation wage as determined by skills, experience, and ability in relation to the current wage obtained. The

length of search should be determined by the rate at which
wage offers at different levels appear in relation to the
individual's reservation wage or aspiration and the cost of
searching. The creation and termination of unemployment
spells in open position systems should be strongly dependent
on individual characteristics.

In closed systems, unemployment processes again will be set
in motion by demand chocks in product markets. However, the
internal labor market firms that are characteristic of
closed position systems do not adjust to product demand
changes by changing wage rates, as in open position systems.
Rather they adjust by changing the quantity produced and lay
off workers in response to lower product demand. Since the
very forces that create closed employment relations make it
costly to lose workers in which firms have made specific
on-the-job training investments, layoffs tend to be initial-
ly short term and with promise of recall. Typically a whole
production unit is laid off. Hence, except for possible
seniority rules imposed by collective agreements, layoffs
should be unrelated to individual characteristics if the
"risk set" is properly specified. The length of the layoff
is again firm rather than individually determined as long as
the recall is in effect.

Quite different hypotheses about the relation between
individuals, firms, and unemployment processes in the two
systems follow from these considerations. For the case of
unemployment durations or the rate of reemployment, these
hypotheses are elaborated in the next section with a special
emphasis on the different mechanisms determining unemploy-
ment durations.

UNEMPLOYMENT DURATIONS IN CLOSED AND OPEN EMPLOYMENT

The typical unemployment spell in open employment systems is produced by a voluntary quit for search. In closed employment systems, the spell should result from a temporary layoff. In open systems, the spell is ended by the acceptance of a wage offer that matches the reservation wage chosen by the individual. In closed systems the spell in ended by the recall. This simple scenario, of course, assumes several things. First, that quits from open employment and layoff from closed employment are the only modes of job separation. Second, that people during unemployment spells do not move from one sector to another.

There is a third mode of job separation in both systems: dismissal. Theoretically, such separations should be frequent in open employment. They should be the exception by definition in closed employment. Empirically they are rare. The standard CPS question about this suggests that only a very small proportion of those unemployed have been dismissed. This may, of course, reflect response bias resulting from reluctance to admit having been fired. It also may result from the unemployment spell being very short after dismissal. They presumably are caused by a discrepancy between a person's performance and the current wage rate. Reemployment opportunities are more ample the lower the wage rate. If the spells are very short they are under represented in cross-sectional counts of spells, because of length bias (Salant, 1977), caused by the selectivity of spells sampled by the cross-sectional survey (Sørensen, 1977b). In longitudinal data, such as those employed here, spells caused by dismissals may be more adequately represented. However, whether they originate in open or closed employment, they should result in search.

It is useful to conceive of the search process in manner
similar to job shift process (Sørensen, 1985). The period
of search is a job spell, the task being the search. The
rate of shift out of a job can in general be argued to be
determined by the current rewards obtained in the job in
relation to the potential rewards where the latter are
determined by the individual's resources as measured by
education, ability, and skills. If only money matters, the
main variables should be the current wage in relation to the
reservation wage. Empirical support for a positive partial
effect of rewards and a positive effect of resources on the
rate of shift is presented for example in Sørensen and Tuma
(1981). For unemployment spells, the rewards of the "job"
are unemployment compensation, welfare, support from other
family members, and leisure. The potential rewards are
those hoped for from the search and again should reflect the
person's skills, experience, and other resources. As in job
shift we should expect that the larger the discrepancy
between these two sets of variables, the more likely it is
that better employment can be found and the higher the rate
of reemployment.

If job offers appear at a constant rate, we should expect
time constant rates of reemployment if the individual's
aspiration for better employment or his reservation wage
remains constant. Time constant rates means that the rate
at which spells end should be the same regardless of how
long the person has already been searching. Departures from
the assumption of a constant rate of job offers would be
caused by cyclical changes. They are difficult to model
with the data used here. Given the quite short periods
under consideration, it seems reasonable to ignore such
changes in the employment offer distributions. However,
there are presumably differences by locale in the distribu-
tion of job offers. Such variation is a source of unmea-
sured heterogeneity.

It may not be reasonable to assume that a person's
aspiration level or reservation wage remains unchanged
indefinitely during a search. The longer the search has
already gone on without success, the more discouraged the
individual. This should lead to a lowering of the aspiration
level and an increased probability of reemployment. One
should therefore expect positive duration dependence in
spells of unemployment due to search, that is, the rate of
reemployment should increase as the spell progresses
(Lippman and McCall, 1976).

The rate of reemployment after an unemployment spell due to
search should then be a question of the individual
characteristics that determine his aspiration level or
reservation wage, the amount of income available during the
spell, and the duration of the spell already completed.
These hypotheses will be tested below with the emphasis on
establishing the predicted positive duration dependence.

The length of temporary layoffs from closed employment
should, as noted, not depend on individual characteristics,
but be completely determined by the rate of recall. This
rate in turn should vary between firms reflecting their
particular production schedules. This heterogeneity in the
rates of recall when left unmeasured should produce observed
negative time dependency in the rates of reemployment. This
phenomenon of "spurious" time dependency due to unmeasured
heterogeneity is well known and results from the changing
composition of those remaining in the state (here of being
unemployed), since those with high rates leave first.

Temporary layoff is an ambiguous state. What starts as a
temporary layoff may become an indefinite layoff. Such a
shift would result in genuine negative time dependency in
the rate of reemployment if the unemployed remains in the
closed employment sector. The job structures of closed

employment are such that new hires tend to take place only
at the bottom levels of job ladders because of the need to
maintain promotion schedules and training arrangements. The
unemployed from the closed sector tend to have firm specific
skills and experiences that are less employable in other
firms. Thus as the chances for recall diminishes, the rate
of reemployment should also decrease.

The negative time dependence in closed employment sectors
may be argued to be eventually changed by a shift into the
open employment sector and search in a competitive labor
market. The prediction then should be that the positive time
dependency eventually becomes positive as the rate of
reemployment also becomes dependent on individual charac-
teristics.

It is hypothesized that there are sources of both positive
and negative true time dependency that should be
identifiable by labor market structures. However, there
will also be unmeasured heterogeneity that would show up as
negative time dependency regardless of the labor market
structure the individual is exposed to. The predicted
differences therefore may only show up as more or less
negative time dependency.

These hypotheses are tested below using data on employment
spells from the Panel Study of Income Dynamics (PSID)[3] In
the present analysis no direct measures of labor market
structures are introduced. Such measures relying on occupa-
tional and industry variables are being constructed in
current research. Here, I shall rely instead on the indi-
vidual characteristics of race and labor force experience as
indicators of labor market experience, assuming that blacks
and inexperienced individuals are more likely to be located
in the open employment sector. Similar information on
whether or not the individual returned to the same employer

after the unemployment spell will be used to indicate closed
employment.

DATA AND METHODS

The present analysis uses the 1982 information about spells
of unemployment. Respondents in the 1982 wave were asked
about their current employment status. If they answered
unemployed or temporarily laid off they were asked a series
of questions about the amount of unemployment, unemployment
compensation, etc. If employed they were asked if they had
been unemployed in 1981, and about the timing of the spells.
The present analysis used information about the most recent
spell including spells in progress at the time of interview
(treated as censored) for male heads of households. This
produced 579 spells from the 3,384 male heads of households
who responded to the survey in 1981.

PSID contains a variety of information about the
unemployment experiences of male heads. From this informa-
tion a variety of variables were constructed. They are
described in Table 1.

Table 1 Here

A few comments about some of these variables are in order.
The variable CHNP is a measure of unemployment compensation
received in 1981 dividing the total amount of compensation
received by the number of hours reported to be unemployed.
This is an average figure that does not reflect unemployment
compensation received before 1981. Less than half received
any unemployment compensation. The dummy variable CHUN

measures this. The variable HAHE is the respondent's
average wage rate in 1981. It is here used to measure the
respondent's resources in the manner suggested by the
conception of the search process as a job shift process.
This wage rate will be partly endogenous to the unemployment
process, and a better measure would be the respondent's
predicted wage rate (Atkinson, Gomulka, and Micklewright,
1984). An attempt was made to obtain predicted wage rates
using variables (education and experience) available in the
present analysis to estimate a wage equation for the whole
sample of male heads. The resulting predicted wage rates
for the unemployed did not have a significant effect on the
reemployment rates and the apparent unreliability bias other
coefficients. Hence, despite the possible endogeniety, the
actual wage rates were used.

In some spells, 32 to be exact, the unemployed lost
unemployment compensation during the spell. This was taken
into account in computing the hourly unemployment compensa-
tion. In addition, the unemployment compensation variable
is treated as a time varying co-variate in the analysis.
Finally, the time dependent co-variate NBF was constructed
to indicate that unemployment compensation was lost.

The measure of the duration of the unemployment spell was
obtained using information about the duration in weeks of
the most recent spell in 1981 or the number of weeks of
unemployment completed for those unemployed at the time of
interview. As shown in Table 1, there is a considerable
amount of censoring present. Given the censoring, the most
informative description of these data is provided by the
Kaplan-Meier estimate of the survivor function providing an
estimate in the presence of censoring of F(t), that is the
proportion remaining unemployed by time t. Selected values
of the Survivor function are presented in Table 2.

Table 2 Here

While there indeed are a number of short spells of
unemployment, the median is more than 12 weeks. There are a
fairly substantial number of long spells -- 20% are
estimated to last more than a year.

From this information event history models for the rate of
reemployment can be estimated with variables listed in Table
1 as co-variates. A fully parameterized model such as the
Weibull or the Gompertz would seem desirable. These models
would allow the estimation of a time dependency parameter
from which the direction of the duration dependence could be
inferred. However, available software did not permit the
estimation of the interaction between co-variates and the
time dependency[4]. Because the statistical significance of
such interactions is of interest here, the discrete time
approximation to the continuous time models proposed by
Allison (1982) was chosen. This approach also easily allows
for the incorporation of time varying co-variates.

Allison shows that a discrete time approximation allows for
a very flexible formulation of most event history models.
Using a slight generalization of the framework proposed by
Allison, general models of the sort:

$$p_i(t) = f(A_i(t) + BX_i) \qquad (1)$$

can be estimated with the proper specification of f. Here
$p_i(t)$ is the probability that an event takes place in period
t for individual i, $A_i(t)$ is a vector of constants denoting
the time dependency for individual i, and BX_i is a

set of co-variates characterizing individual i. A particular attractive specification of the function for $p_i(t)$ is:

$$p_i(t) = 1 - \exp[- \exp(A_i(t) + BX_i)] \qquad (2)$$

because it will provide estimates that approximate those that would have been obtained had the proportional hazards or the Cox model been estimated for the underlying continuous time model. In particular, these estimates will not be dependent on the length of the time interval. This specification can be estimated using the complementary log-log "link" specification in GLIM.

If now $A_i(t)$ is further specified as:

$$A_i(t) = a_o + a\log t_i \qquad (3)$$

we obtain an approximation to the Weibull model. Interactions may here be tested using interaction terms involving logt and the co-variates.

The discrete time approximation demands that the data with the unit of analysis being spells be converted to observations on each time unit for each individual. This produces an enormous amount of information -- here about 10,000 observations. Since the event is quite rare and it was necessary to reduce the sample size considerably because of hardware limitations, the time unit was redefined as a two week period and a random 30% sample from these observations were selected. This gave a sample of 2452 observations for the present analysis.

RESULTS

As noted above, I take the average wage rate as a measure of
the respondent's resources. In a search theory interpreta-
tion it should then be assumed that the reservation wage is
a linear function of this quantity. The income provided in
the spell is measured by the unemployment compensation
received. With the conception of search as a job shift and
from search theory it should then be expected that in the
open employment sector the rate of reemployment would be
positively related to the wage rate and negatively related
to the amount of compensation received.

Changes in the aspiration level are assumed to depend on the
time already spent searching so that an increase in the rate
of reemployment will be observed. This should be more
pronounced the more likely it is that the respondent is
located in the open sector. It seems plausible to assume
that inexperienced workers are more likely to be located in
the open sector and to be engaging in search while unem-
ployed. Hence, I expect that younger workers have more
positive time dependency than more experienced respondents.
Similarly, minority workers can be assumed to be more likely
to be in the open sector, so that the same interaction will
be observed for race.

Results of the complementary log-log estimation of various
models are shown in Table 3.

Table 3 Here

The results are very much conforming to predictions. First,
note that the model that does not include the interaction

terms show no significant effect of experience. The model
that includes the interaction terms changes the estimates of
the effect of both race and of experience considerably,
making the main effect of experience highly significant and
more than doubling the effect of race. Both the interac-
tions with time are significant. Whites have more negative
time dependency than blacks and inexperienced workers have
more positive time dependency than experienced workers.
Thus, young blacks have the most positive time dependency
while experienced whites have the most negative.

The effect of amount of unemployment compensation is
significantly negative. The more compensation received, the
longer the spell, if respondents receive any compensation at
all. It is of considerable interest to note that the spline
function fitted demonstrates that there appears to be a
positive effect on the rate of reemployment of being covered
by unemployment compensation. Thus for low levels of
compensation there may be a higher rate of reemployment than
for those that do not receive any compensation at all. This
complicated relation between compensation and rates of
reemployment suggests why there appears to be continuing
controversy about the impact on compensation on duration of
unemployment (Atkinson, Gomulka, and Micklewright, 1984).

The effect of the average wage rate received in employment
is positive, as predicted when this variable is interpreted
as a measure of the persons resources. Whites and
experienced workers have much higher rates of reemployment
than blacks and inexperienced workers. Thus the frequent
finding that young or black workers have short spells is
largely a result of the positive time dependency. If these
workers were in the same employment sectors as whites and
experienced workers their spells would be considerably
longer.

The third model in Table 3 shows the effect of adding a squared term in logt to the model. As noted above one might expect that the time dependency becomes positive after very long spells. This prediction is not borne out -- the effect is non-significant, indicating no pronounced curvelinearity.

The fourth model now adds the measure of having lost unemployment benefits. As expected there is a significant positive effect on the rate of reemployment.

It is of interest to directly identify the spells most likely to occur in the closed sector where dependency on measured characteristics is hypothesized to be less. PSID asks if the respondents return to the same employer after the spell. These spells are then likely to be layoffs in the closed sector. This information can be used to estimate a competing risk model for the two outcomes of the spell: returning to the same employer versus not returning to the same employer. The results of estimating the main model of Table 3 for these two outcome are presented in the first panel of Table 4.

Table 4 Here

As predicted, few variables show a significant effect on reemployment rates for those who should be more likely to be in the closed sector. Only experience and time have significant effects. As predicted there is more negative time dependency for these spells. Those spells not result- ing in return to the same employer in contrast have no significant time dependence and a strong effect of race. In particular it should be noted that the effects of unemploy- ment compensation and wage rates are substantially smaller. This presumably shows that search is less likely to govern

the unemployment spell in this more closed sector. However, the second panel of Table 4 also shows that the effect of experience does not show up either in the more "open" sector where lay off is not indicated. Presumably, in the more open sector, experience as such does not discriminate among different employment processes.

Return to the same employer, of course, could also take place in the open sector. Another definition can be obtained by grouping together spells covered by unemployment compensation the closed sector. Correspondingly, those spells where compensation was not received are classified as the more irregular or open sector. Since we here cannot use amount of compensation as a variable, only experience and race are included as independent variables. The results are presented in Table 5.

Table 5 Here

In the more regular, closed sector, none of the independent variables have an effect. In the more open, irregular sector, race shows very strong effects, while as before with return to the same employer, experience makes little difference. Note also that the rates of reemployment are very much higher in the more closed sector than in the more open sector.

CONCLUSION

This paper has shown that unemployment processes are quite heterogeneous according to employment sector. Using indirect indicators to identify sectors it is demonstrated that in the more closed employment sector of the labor market, the duration of unemployment processes is less

influenced by individual level characteristics, including the unemployment compensation received, and that remaining heterogeneity results in apparent negative time dependency. In the more open sector, the unemployment process is more dependent on individual decision making as predicted by search theory. In particular, individual change over the unemployment process results in positive time dependency for those most likely to be located in the open sector, inexperienced workers and minorities.

Of particular policy interest is the quite complicated mechanism detected for how unemployment compensation influences duration of unemployment. Increased compensation does increase unemployment spells, as predicted by search theory. However, those covered by compensation appears to be more likely to be located in the more closed sector where unemployment spells tend to be shorter and, in fact, less dependent on individual characteristics, including the amount of compensation. Without disaggregation, the effects of unemployment compensation will, therefore, depend on the distribution of spells among sectors and results from different samples will be inconsistent with each other -- as the controversy about this amply shows.

The next step is, of course, to rely on direct, rather than indirect measures of unemployment sectors. This is done in ongoing research where the life cycle variation in these processes also occupies attention.

NOTES

1. This is the US definition and the one recommended by the ILO. Many European countries also use measures based on counts of the insured instead of, or in addition to, the survey measure.

2. Quantitative sociological research on unemployment events is in general very rare. For one of the few recent exceptions, see DiPrete (1981).

3. For a description of the study see for example Morgan and Duncan (1983).

4. Such interactions can be estimated for the Gompertz model in RATE (Tuma, 1979).

5. Experience is dichotomized as it can be shown that the effect is clearly non-linear.

REFERENCES

Allison, Paul D. 1982. Discrete-time methods for the analy-
    sis of event histories. pp. 61-98 in S. Leinhardt (ed.)
    Sociological Methodology 1982. San Francisco:
    Jossey-Bass

Atkinson, A. B., J. Gomulka, and J. Micklewright. 1984.
    Unemployment benefit, duration and incentives in
    Britain. Journal of Public Economics 23:3-26

Azariadis, Costa. 1975. Implicit contracts and underemploy-
    ment equilibria. Journal of Political Economy.
    83:11183-1202

Clark, Kim and Lawrence H. Summers. 1979. Labor Dynamics and
    unemployment: a reconsideration. Brookings Papers on
    Economic Activity 1:1979, pp. 13-60.

DiPrete, T. A. 1981. Unemployment over the life cycle:
    racial differences and the effect of changing economic
    conditions. American Journal of Sociology 87:286-307

Doeringer, P.B. and M. Piore. 1971. Internal Labor Markets
    and Manpower Analysis. Lexington: D.C. Heath

Feldstein, Martin A. 1975. The importance of temporary
    layoffs: an empirical analysis. Brookings Papers on
    Economic Activity. 3:1975, pp. 725-44.

Hall, Robert E. 1970. Why is unemployment so high at full
    employment. Brookings Papers on Economic Activity
    3:1970, pp. 369-402.

Lippman, S. and J.J. McCall. 1976. The economics of job
    search: a survey. Economic Inquiry. 14:155-189.

Morgan, J.J. and G.J. Duncan. 1983. Five Thousand American
    Families. Vol. 10. Ann Arbor: University of Michigan.
    Institute for Social Research.

Salant, Stephen W. 1977. Search theory and duration data: a
    theory of sorts. Quarterly Journal of Economics. 91:
    39-57.

Schervish, P. 1983. The Structural determinants of
    Unemployment. New York: Academic Press.

Sørensen, Aa. B. 1977a. The Structure of inequality and the
    process of attainment. American Sociological Review
    42:965-978

Sørensen, Aa. B. 1977b. Estimating rates from retrospective
     questions. In D. Heise (ed.) Sociological Methodology
     1977. San Francisco: Jossey Bass.

Sørensen, Aa. B. 1979. A model and a metric for the process
     of intragenerational status attainment. American
     Journal of Sociology 85:361-384

Sørensen, Aa. B. 1983. Processes of allocation to open and
     closed position in social structure. Zeitschrift fuer
     Soziologie 12:203-224

Sørensen, Aa. B. 1984. Interpreting time dependency in
     career processes. Pp. 89-122 in A. Diekmann and P.
     Mitter (eds.) Stochastic modelling of Social Processes.
     New York: Academic Press.

Sørensen, Aa. B. 1985. Employment relations and employment
     processes. Paper presented at the Conference on
     Unemployment organized by the Danish Social Science
     Research Council, September 26-28, Ronne, Denmark.

Sørensen, Aa. B. and N. B. Tuma. 1981. Labor market
     structures and rates of job shifts. In D. Treiman (ed.)
     Research in Social Stratification and Mobility.
     1:361-84. New York: JAI Press.

Tuma, N. B. 1979. Invoking RATE (2nd ed.). Menlo Park, CA:
     SRI International.

Weber, M. 1968. Economy and Society. New York: Bedminster
     Press.

Williamson, O. E. 1975. Markets and Hierarchies: Analysis
     and Antitrust Implications. New York: Free Press

Table 1. Variables and Their Means and Definitions.

Name	Mean		Definition
CHUN	.45		1 if received unemployment compensation. 0 otherwise.
CHNP	$2.81	(CHUN=1)	Hourly unemployment compensation in 1981.
HAHE	$7.39		Average hourly wage in 1981.
NBF	.05		1 if benefits lost. 0 otherwise
SAEP	.52	(CEN=1)	Return to same employer
MFTW	.54		1 if full time labor force experience less than 6 years. 0 otherwise.
MWHI	.55		1 if respondent is white. 0 otherwise.
CEN	.66		1 if spell not censored. 0 otherwise.
TIME	16.44	(CEN=1)	Number of weeks unememployed
	22.42	(CEN=0)	
LOGT			Logarithm of TIME.

Table 2. Selected Values of the Kaplan-Meier Estimate of the
Survivor Function, F(t).

Week (t)	F(t)	n(t)
0	1.00	579
1	.94	532
2	.87	488
4	.77	412
8	.67	347
12	.57	276
26	.34	109
39	.24	59
52	.21	24

Table 3. Complementary Log-Log Estimation of Models for Rate of Re-Employment

Independent Variables:

A.

	B	s.e	B	s.e.
Constant	-2.033	.247	-2.179	.412
CHUN	.392	.216	.394	.218
CHNP	-.127	.063	-.130	.064
HAHE	.031	.013	.034	.013
MFTW	.031	.153	-.786	.347
MWHI	.591	.161	1.372	.393
LOGT	-.480	.077	-.453	.179
LOGT*MFTW			.421	.163
LOGT*MWHI			-.375	.172

-2*Log-likelihood   1232.4   1217.1

B.

	B	s.e.	B	s.e.
Constant	-2.116	.508	-2.167	.409
CHUN	.394	.218	.110	.221
CHNP	-.130	.064	-.046	.062
HAHE	.033	.013	.027	.013
MFTW	-.776	.346	-.746	.343
MWHI	1.360	.394	1.460	.390
LOGT	-.530	.406	-.484	.176
LOGT*MFTW	.417	.163	.417	.161
LOGT*MWHI	-.369	.173	-.389	.169
LOGT*LOGT	.019	.088		
NBF			1.713	.316

-2*Log-Likelihood   1217.0   1195.5

N = 2266

Table 4. Models for Rates of Re-employment for Spells ending in Return to Same Employer and No Return to Same Employer

	Same Employer		Not Same Employer	
	B	s.e.	B	s.e.
Constant	-1.603	.487	-5.370	.783
CHUN	.532	.303	.390	.322
CHNP	-.076	.078	-.253	.114
HAHE	.032	.018	.028	.021
MFTW	-1.268	.463	.841	.600
MWHI	0.621	.482	2.428	.680
LOGT	-1.069	.257	.509	.287
LOGT*MFTW	.415	.250	-.065	.242
LOGT*MWHI	-.036	.261	-.732	.256
-2*Log-Likelihood	686.9		744.21	
N	2262			

Table 5. Models for Rates of Re-employment for Spells in the "Closed" and the "Open" Sector

Sector	"Closed"		"Open"	
	B	s.e.	B	s.e.
Constant	-1.308	.933	-2.425	.594
MFTW	-.831	.826	-.361	.488
MWHI	.834	.966	1.893	.551
LOGT	-.785	.484	-.493	.273
LOGT*MFTW	.628	.419	.398	.243
LOGT*MWHI	-.134	.492	-.455	.243
-2*Log-Likelihood	223.55		602.63	
N	376		1311	

Note: "Closed" sector is defined as spells covered by unemployment compensation of more than $3.00. Open is defined as not covered by unemployment compensation.

Unemployment duration as a function of individual characteristics and economic trends

Reinhard Hujer and Hilmar Schneider

1. Introduction

Although unemployment represents one of the most urgent problems of economic and social policy since over a decade, the existing variety of theoretical approaches is confusing and empirically insufficiently tested (Buttler 1984). Accordingly, the adequate political measures to be taken evoke controverse. The following paper does not claim to be able to do away with these deficiencies. However, it does point to aspects which have been somewhat neglected in previous discussions.

The ongoing vocational crisis in the Federal Republic developed in two marked shifts, closely related to the oil price shocks, which occurred at the end of 1973 and again prolonged over a period from 1978 to 1981. On both occasions, the crisis led to a period of flat growth because the increasing prices of resources could not be compensated

despite growing productivity.[*] Unfortunately, the data presented in this paper only covers the period from 1970 to 1979. Therefore, the following statements are restricted to consequences of the first oil-crisis. In 1974 and 1975 the unemployment rate increased rapidly. Annual man hours worked decreased as well. At the same time the working population declined although this trend later reversed because of a decreased effective working time, i.e. the average annual amount of man hours. The actual increase in unemployment during 1974 and 1975 would have been even greater, if the number of persons able to work (i.e. the number of persons between 15 and 65) had not stagnated at the same time.

Analyzing unemployment stock, a growth of unemployment figures can have two causes: A relative growth of inflow or an increase of the mean unemployment duration (Egle 1984). Nevertheless, most of the empirical models concerned with unemployment are restricted to analysis of stock itself (ibid.). Among the few exceptions for the Federal Republic are the works of Freiburghaus (1978) and Lempert-Helm (1985). Bailey and Parikh (1985) conclude in a more recent study for Great Britain that the increase of unemployment during the seventies was mainly the result of an increase of mean unemployment duration. However, all of these works are based on aggregate data. The lack of individual data concerning unemployment duration is one of the most important limitations on these studies.

Although Büchtemann's paper (1985) is based on individual data, it is more descriptive than model oriented. In addition to the data presented here concerning the seventies, data from the Socio-economic Panel (see Hujer/Schneider, 1986) should improve this situation.

---

[*] Monatsberichte der Deutschen Bundesbank, April 1978, pp. 22 f.; Monatsberichte der deutschen Bundesbank, April 1981, pp. 13 f.; Friedrich/Ronning (1985)

One of the questions that may be answered with our analysis is, whether the overall ascertained increase of unemployment duration has to be classified as real or as spurious. A real increase would be present if it took place on the aggregate level as well as on the individual level. A spurious increase would result if the individual-specific unemployment duration remained constant throughout the observation period while the probability of inflows to unemployment increased for persons with higher individual-specific unemployment duration. When the former case holds true, a significant influence of labour market indicators should be recognized. This, however, may not happen in the latter case.

Beyond this, individual data allow a differentiation between inflows and outflows to and from unemployment by different prior and destination states. This possibility is represented here by two submodels that separate outflows from unemployment into two destination states: 'return to occupation' and 'withdrawal to non-activity'.

Of course, the analysis of stocks and flows touches only the surface of the problem. Micro-economic approaches try to come to a more profound explanation of the labour market (Helberger 1982). The present paper, however, must be restricted to analysis of part of the question.

Job search theorists (Lippman/McCall 1976) argue that the decision of an unemployed person to take up employment depends on two major components: Firstly, a job searcher individually fixes his optimal acceptance wage. The acceptance wage is inversely related to search and opportunity costs. The second component is the wage that is actually offered for a job. A positive decision to occupy a job will take place, when the offered wage exceeds the acceptance wage.

It follows from these arguments that a decline in opportunity costs, induced primarily by an increase in unemployment

compensation, increases search duration. An increase in the level of unemployment results. Job search theory thus led to discussion of whether the level of unemployment compensation could be used to control the level of unemployment.

Although the assumptions differ, advocates of contract theory come to similar conclusions regarding the level of unemployment compensation and its influence on the level of employment. From their point of view, unemployment may increase because firms temporarily dismiss employees during periods of capacity underutilization (Feldstein 1975). The lower the related losses for the two contracting parties (i.e. employer and employee), the sooner such behaviour occurs. The applicability of this approach, however, is restricted due to the fact that temporary unemployment constitutes only a part of total unemployment.

Empirically, the effect of the level of unemployment benefits has been analysed most extensively in the U.S. (see Hamermesh 1977 with a summary of then current results). A more recent study of the U.S. by Moffitt (1985) is methodologically similar to this work. His results largely confirm the above conclusions. However, one has to take into account that job offers that call for a decision on the part of the job searcher may occur only rarely during a vocational crisis. A reduction of unemployment compensation will therefore have little effect upon unemployment duration, while it will surely increase the problems of the unemployed.

Finally, human capital theorists argue that unemployment may be explained by discrepancies between individual qualifications and certification requirements of labour demand (Helberger 1982). Although this argument may hold true outside of vocational crises, it seems to be too simplistic otherwise. It does not appropriately reflect that even high

qualifications may not prevent unemployment during a recession.

In general, a problem of these micro-economic labour market theories is their failure to adequately consider global developments in labour demand. It is our objective to overcome this disadvantage by formulating models which incorporate global labour demand as well as individual labour supply components.

2. Rate Models: Basic Concepts and Estimation

Rate models can be used to analyse the problems of duration of unemployment, since they are suitable for modelling the change in qualitative variables by means of event-history analysis (Tuma/Hannan 1984, Lawless 1982, Arminger 1984, Diekmann/Mitter 1984, Cox/Oakes 1984). The appropriate mathematical specification of the dynamic process generating the data is that of a continuous-time, discrete-state Markov or Semi-Markov process. The Markov assumption means that the process of leaving a presently occupied state is dependent on the immediately preceeding state of origin. For example, the birth of a child is an important cause for change from unemployment to non-activity. This assumption may at times be problematic, i.e. when the relevant factors are not immediately preceeding the state of interest (see Figure 1 for an example). As will be seen later, we allow for such violations of the Markov assumption by explicitly modelling earlier events, such as child-bearing.

The structure of rate models can be described by the following basic relations. If $Z(t)$ is a random variable denoting the state of the dependent categorial variable at time t and $q(t_0, t) = (q_{ik}(t_0, t))$ is the matrix of transition probabilities of moving from state j to state k in the time interval $[t_0, t]$, then we may define the transition rate $r_{jk}(t)$ as:

| Figure 1 | hypothetical transitions from occupation to non-activity |
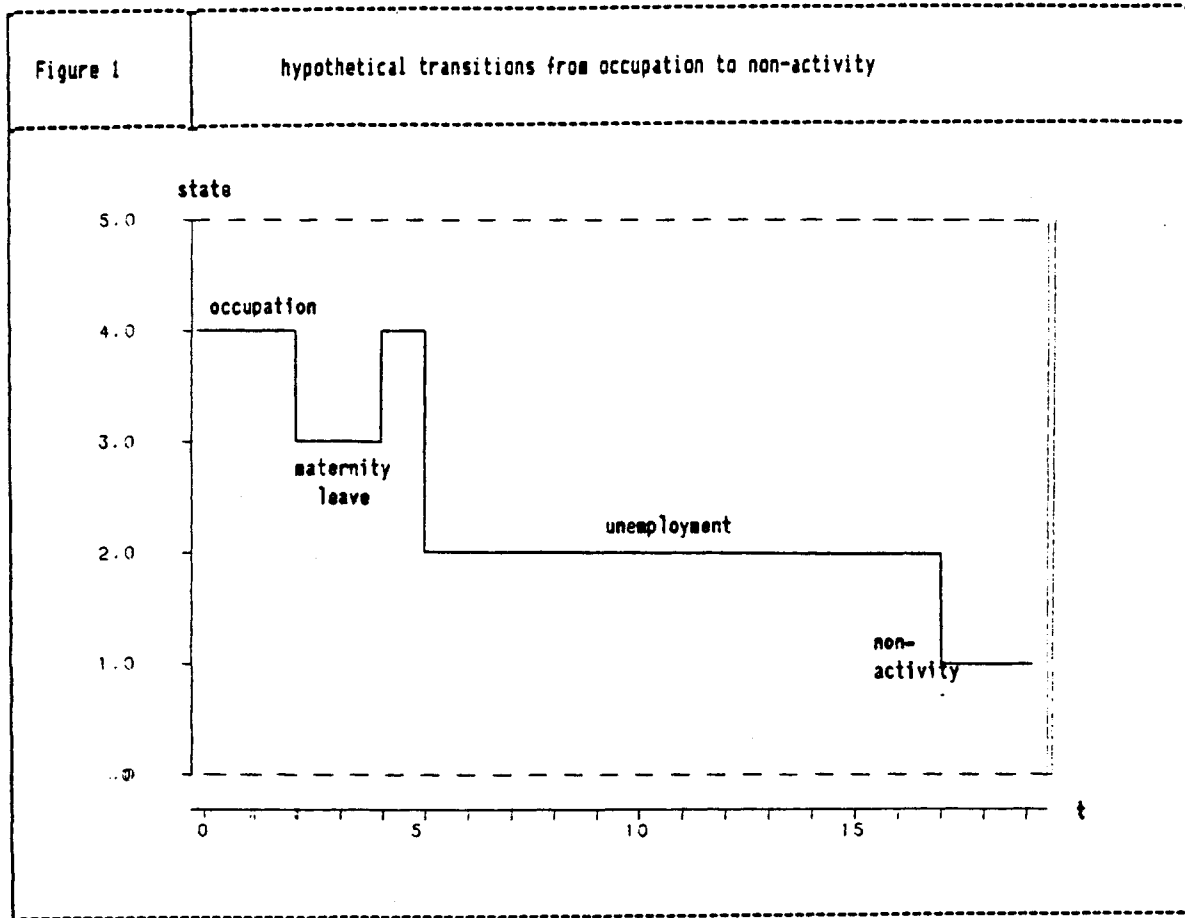
$$(1) \qquad r_{jk}(t) = \lim_{\Delta t \to 0} \frac{q_{jk}(t,t+\Delta t)}{\Delta t} \geq 0 \mid j \neq k$$

The rate of leaving state j (hazard rate) is expressed for a multi state model by:

$$(2) \qquad r_j(t) = \sum_{j=1}^{K} r_{jk}(t) \qquad \text{with } k \neq j, \; k = \text{number of states}$$

The rates $r_{jk}(t)$ are the essential parameters of a continuous-time, discrete-state semi-Markov process. For the purpose of estimation we need several distributional functions. Denoting the probability of leaving state j - occupied at time t_0 - before or at time t with $F_j(t|t_0)$, the relation between F_j and r_j follows:

$$(3) \qquad F_j(t|t_0) = 1 - \left[\exp\left(-\int_{t_0}^{t} r_j(\tau)\,d\tau\right)\right]$$

The survivor function G_j is then:

$$(4) \qquad G_j(t|t_0) = 1 - F_j(t|t_0)$$

The density function f_j has the following form:

$$(5) \qquad f_j(t|t_0) = \frac{dF_j(t|t_0)}{dt} = r_j(t)\, G_j(t|t_0)$$

Finally, the moments of duration T are expressed by:

$$(6) \qquad E(T_j^m) = \int_{0}^{\infty} \tau^m\, f_j(\tau)\, d\tau$$

The most important purpose of rate modelling is to explain a dynamic process by specification of a hazard function. On one hand it is necessary to model time dependence (Semi-Markov process), on the other hand we are interested in considering the influence of exogenous variables (covariates) on the rates:

$$(7) \qquad r_{jk}(t) = r_{0jk}(t)\, A_{jk}(t) \qquad \text{with } g(A_{jk}(t)) = \underline{x}'_{jk}(t)\, \underline{\beta}_{jk}$$

$r_{0jk}(t)$ is a time dependent base line rate with a mathematical specification and $g(A_{jk}(t))$ is the link function. In empirical analysis we used a log-linear model $g(A_{jk}(t)) = $

$\ln(A_{jk}(t))$. The simplest parametric model of time dependence is an exponential model with constant rates:

$$(8) \qquad r_j(t) = r_j$$

In this special case the duration can be interpreted in a simple manner:

$$(9) \qquad E(T_j) = \frac{1}{r_j}$$

Considering two covariates x_1 and x_2, the model for the duration in state 0 has the following form:

$$(10) \qquad E(T_0) = \frac{1}{r_{01} + r_{02}}$$

$$\text{with } r_{01} = \alpha_{0(01)} \; \alpha_{1(01)}^{x_1} \; \alpha_{2(01)}^{x_2}$$

$$r_{02} = \alpha_{0(02)} \; \alpha_{1(02)}^{x_1} \; \alpha_{2(02)}^{x_2}$$

$$\alpha = e^{\beta}$$

If $r_{10} = r_{20} = r_{12} = r_{21} = 0$, we consider the special case of competing risks. It follows, that the time of event of individual i is then $t_i = \min(T_1, \ldots, T_j, \ldots, T_k)$.

To determine the transition rates we can use maximum likelihood estimation. Considering a two-state model with $i = 1, \ldots, n$ independent observations, given for each i either a time of event or a time of censoring, we can derive the likelihood function as:

$$(11) \qquad L = \prod_{i=1}^{n} f_i(t_i)^{c_i} \, G_i(t_i)^{1-c_i} = \prod_{i=1}^{n} r_i(t_i)^{c_i} \, G_i(t_i)$$

$$\text{with } c_i = \begin{cases} 1, & \text{if an event happens to person } i \text{ during the period of observation} \\ 0, & \text{if observations are censored} \end{cases}$$

This basic approach can be extended to repeated events $h = 1, 2, \ldots, \infty$. In this case the likelihood function is:

$$(12) \qquad L = \prod_{h=1}^{\infty} \prod_{i=1}^{n} r_i(t_{ih}|t_{i(h-1)})^{c_{ih}} G_i(t_{ih}|t_{i(h-1)})$$

Considering the model of competing risks, it follows (Lawless 1982, p. 475 ff.):

$$(13) \qquad L = \prod_{i=1}^{n} f_{ij}(t_i)^{c_i} G_{ij}(t_i)^{1-c_i} \prod_{k \neq j}^{K} G_{ik}(t_i)$$

With $f_{ij}(t_i) = r_{ij}(t_i) G_{ij}(t_i)$ we can reformulate Eq. (13):

$$(14) \qquad L = \prod_{j=1}^{K} \left[\prod_{i=1}^{n} r_{ij}(t_i)^{c_i} G_{ij}(t_i) \right]$$

Time dependency can also be incorporated in considering time-dependent covariates. Kalbfleisch and Prentice (1980) distinguished between external and internal covariates. External time-dependent covariates are exogenous and independent of the dynamic process, while internal covariates are a function of the process and the individual under study. External covariates are for example the level of unemployment compensation or labour market indicators; examples of internal covariates are indicators for mental health. In many cases however, it is not possible to differentiate exactly between internal and external covariates.

A simple model which takes such time-dependent effects into account includes an explicit function of time $x(t) = f(t)$. This approach can be used in a regression model. A second approach consists of a period model (Tuma/Hannan 1984),

which is suitable for modelling time-dependent covariates changing in discrete time. In this model the observation period is divided into p subperiods. If τ_{p-1} denotes the beginning and τ_p the end of the period p, the specific transition rate of the period $r_{ijkp}(t)$ is then:

$$(15) \quad \ln(r_{ijkp}(t)) = \underline{x}'_{ip} \, \underline{\beta}_{jkp} \quad \text{for all t with } \tau_{p-1} \leq t < \tau_p$$

The survivor function $G_{ijk}(t)$ of the observation period $p=1,2,\ldots,p$ can be expressed by:

$$(16) \quad G_{ijk}(t) = \exp\left(-\sum_{p=1}^{p} \int_{\tau_{p-1}}^{\tau_p} r_{ijkp}(\tau) \, d\tau\right)$$

The specification of the periods is dependent on special individual characteristics or on the time intervals between observations (panel data). The practical implementation of this concept in the data set is accomplished by splitting a single spell into several subspells, according to the number of changing points for covariate values. An example of this procedure is shown in Table 1. We chose a spell with a length of 17 units, terminated by an event at t=17. We considered two time-dependent covariates, unemployment rate and entitlement to unemployment compensation.

3. The Database

The present analysis is based on individual data from the Lebenslagen-Studie (LeLa) of the Sonderforschungsbereich 3. The main survey was carried out between October 1980 and April 1981. The parent population consisted of the total of all German citizens aged 15 to 60 and living in private households in the Federal Republic of Germany and West-Berlin. The survey sample was drawn at random, stratified by community size. As a special feature of the sample design,

Table 1	example for spell-splitting with time dependent covariates				
Start- time	Ending- time	Censoring status	Entitle- ment to Arbeits- losengeld	End of claim reached	unemploy- ment rate
0	3	0	1	0	4.6
3	10	0	1	0	4.5
10	12	0	1	1	4.5
12	15	0	0	0	4.5
15	17	1	0	0	4.3

the spouses of married target persons were interviewed as well. The total sample comprises 9535 completed interviews.

In addition to the survey data, a subsample of 4165 cases contains data from the officially registered social credit accounts. The latter were made available by the social insurance carriers with the consent of the survey participants. These data allow a retrospective time location of unemployment spells. The data can be viewed as valid since they form the basis for the calculation of pension claims. The period of 1970-1979 contains a total of 851 spells, of which 509 are transitions to employment, 293 are transitions to non-activity, while the rest are right-censored, i.e. these subjects did not change states during the period in question.

The social credit account sample is not fully representative for the economically active population of the Federal Republic, but this should not seriously affect multivariate modelling, as long as there is no misspecification (Hausman 1978). The most relevant deficiencies are that foreigners and retiring people are underrepresented. Foreigners averaged about 7% of the overall population during the observation period but were only drawn into the sample as the

spouse of a German target person. For the same reason the sample is retrospectively biased, because persons who reached age 60 throughout the seventies were not directly addressed by the sample design. As an important consequence, transitions from unemployment to pension cannot be observed. It is known from other work that older unemployed persons often make this transition (Büchtemann/Infratest Sozialforschung 1983).

The available data set is suitable for analysis of partial aspects of the present problem. Employment for example, cannot be differentiated by branches; if done, it would have been inaccurately taken into account due to the restricted sample size. Effects of occupational training were not estimated due to the absence of information in the data.

The following summary illustrates the character of the data. Because of the described selectivity and representativity problems, differences between our data and the officially available data should not be overvalued. In Figure 2, spell durations generated from the sample are compared with the official statistics of the Federal Labour Office (Bundesanstalt für Arbeit). The official data are generated by calculating the average of the previous spell durations for the whole unemployment stock in September of each year.

This procedure leads to an underestimation of the true spell duration. To enable a comparison this concept was reproduced with the Lebenslagen-data.*) The trend of the officially computed unemployment duration rises between 1974 and 1976.Afterwards the trend stabilizes at a level around eight months. In this pattern, we find the above-mentioned structural shift following the oil-crisis, although lagged one year when compared with the rest of the labour market indicators. The observed lag must primarily be ascribed to the computational method used by the Federal Labour Office.

---

*) The curve for the Lebenslagen-data starts at 1973 as too few cases were available in earlier years

| Figure 2 | comparison of mean unemployment duration |
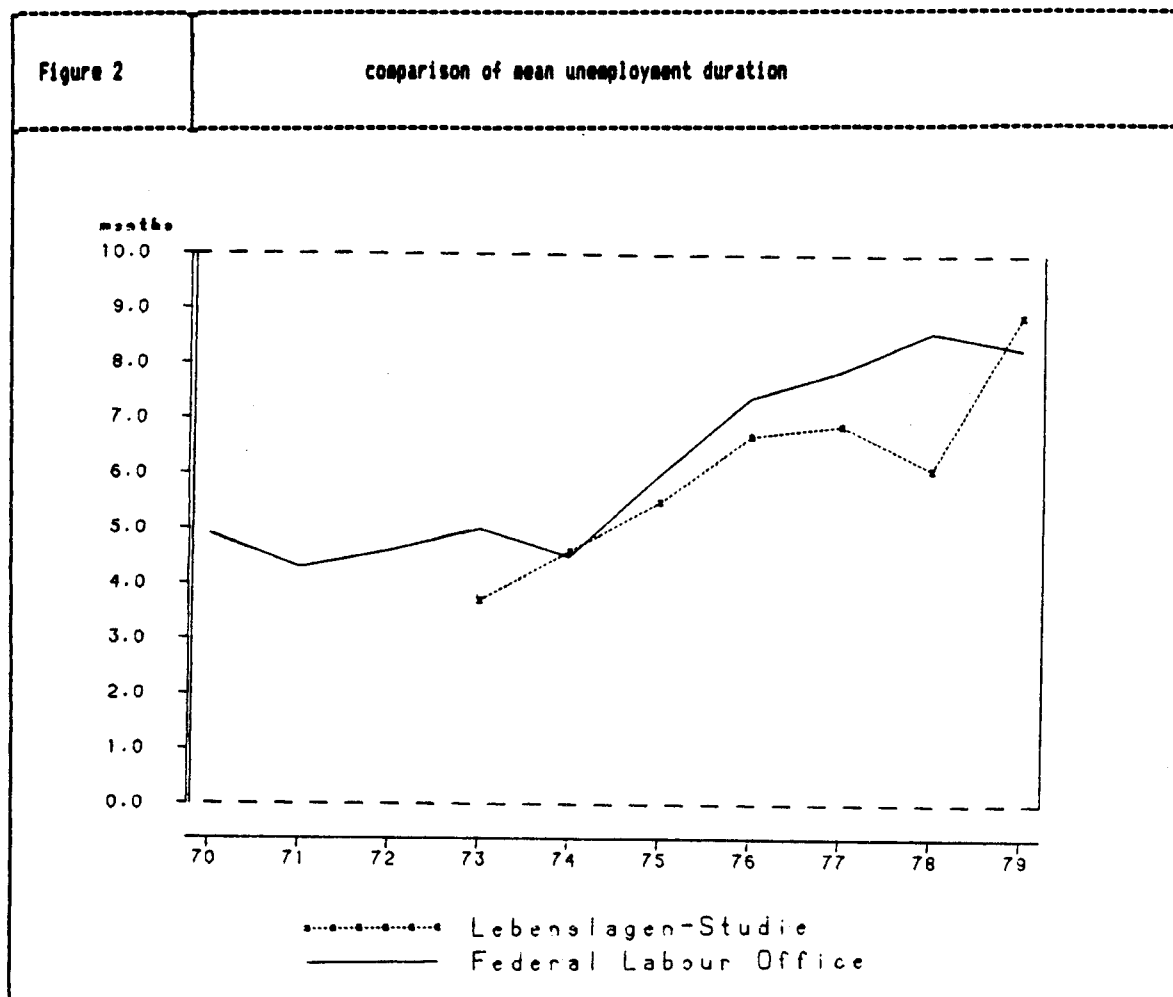
A frequency distribution of the number of spells experienced by the unemployed (Table 2) shows that over two-thirds experienced only one spell throughout the observation period. A relatively small group of persons experience the majority of spells, i.e. about 56%. This agrees with Büchtemann's results (1985) for the nine-year period between 1973 and 1982.

Beyond dating unemployment spells, the data allow estimation of unemployment compensation. Instead of actual income data, we had monthly information about the amount of social credit points at our disposal. These are used to evaluate pension

Table 2	unemployed by number of their spells 1970-1979		
number of spells	n	in % of cases (559)	in % of spells (851)
1	379	67.92	44.56
2	116	20.79	27.26
3	37	6.63	13.04
4	13	2.33	6.11
5	4	0.72	2.35
6	6	1.08	4.23
7	3	0.54	2.47
-----	-----	-----	-----
∑	558		

claims but may also be used to infer personal monthly gross income. The following equation holds approximately:

$$\begin{array}{l}\text{personal} \\ \text{monthly} \\ \text{gross} \\ \text{income}\end{array} = \begin{array}{l}\text{total} \\ \text{annual} \\ \text{gross} \\ \text{income}\end{array} * \frac{\text{social credit points}}{100}$$

This makes it possible to take the number of social credit points as a rough indicator of the real amount of unemployment compensation, although its evaluation is based on the last personal monthly net income. Due to the absence of additional information we used the assumption of a constant relationship between gross and net income.

The German Unemployment Insurance System is predominantly financed through contributions of employed persons excluding civil servants. Eligibility for benefits exists when an employee has contributed for a minimum period within a legally fixed time frame.[*] In the event of unemployment

[*] for details see: Arbeitsförderungsgesetz vom 25.6.1969 and subsequent changes

such persons are first entitled to 'Arbeitslosengeld'. The maximum claim duration for Arbeitslosengeld commonly comprised 12 months during the seventies. This applied to persons, who had been employed for at least 24 months within the last 3 years before unemployment. The level of Arbeitslosengeld amounted to about two-thirds of the last personal monthly net income. After expiration of entitlement to Arbeitslosengeld, an unemployed person is next entitled to 'Arbeitslosenhilfe'. Its maximum claim duration is not legally restricted, but is determined, with a measure of discretion by the respective local Labour Office. The level of Arbeitslosenhilfe depends not merely upon the last personal monthly net income earned, but also takes other sources of income into account, in particular the income of spouses. Thus, Arbeitslosenhilfe may average less than half of the last personal monthly net income. In practice, for the majority of married women, benefit claims expire with the time limit for Arbeitslosengeld.

4. Results

The modelling process took two principal steps: First, we constructed a baseline exponential model on the basis of all available individual variables (Table 3). A significance test of the null hypothesis for single parameters was used to select variables for inclusion into the model. The next step consisted of one-by-one inclusions of the available labour market indicators (Table 4). A simultaneous inclusion of two or more labour market indicators could not be accomplished because of strong multicollinearities. The effects of the inclusion of additional variables on the baseline model can be ignored; thus these changes are not presented separately.

The two tables each contain two submodels according to the presently differentiated destination states. The column 'α' gives the change in the hazard rate when the corresponding

Table 3	baseline exponential model for leaving unemployment[*)]

return to occupation		leaving to non-activity		
α	diff Chi^2	α	diff Chi^2	variable-name
0.0651 **		0.1507 **		constant
1.7370 **	34.25	-		sex (0=woman/1=man)
-		0.6334 **	7.77	educational degree (0=else/1=Realschule)
0.6315 *	6.64	0.4649 **	12.27	educational degree (0=else/1=(Fach-)Abitur)
1.2670 *	6.13	0.6418 **	10.18	occupational status before spell (0=else/ 1=blue-collar worker
-		0.6562 **	8.81	occupational status before spell (0=else/ 1=white-collar worker)
0.7067 *	7.26	0.5004 **	12.76	age (0=else/ 1=50 years and older)
0.4093 **	18.20	-		pregnancy before spell (0=else/1=within last 12 months before Spell)
1.6250 **	20.37	0.5973 **	15.12	entitlement to Arbeitslosen- geld (0=else/1=yes)
-		4.1520 **	45.09	time limit of entitle- ment to Arbeitslosen- geld reached (0=else/1=yes)
-1525.791		-1034.169		log-likelihood
137.09		90.82		global Chi^2
6		7		degrees of freedom

Table 4	extended baseline model for influence of labour market indicators upon the hazard rate for leaving unemployment*)

return to occupation		leaving to non-activity		
α	diff Chi²	α	diff Chi²	variable-name
. . .		. . .		
0.8781 **	13.28	1.0460	0.74	annual unemployment rate
1.0010 **	11.28	0.9998	0.18	annual number of vacancies in thousands
0.9071 **	10.22	1.0440	0.96	annual number of unemployed per vacancy
1.0920	0.34	0.8041	1.20	annual number of employees excluding civil servants in millions
1.0001	0.30	0.9996	1.80	annual number of economically active persons in millions
0.9997 **	7.46	1.0001	0.07	annual number of persons capable of gainful activity (PCGA) in millions
1.0001 **	13.42	0.9999	1.04	annual number of man hours worked in billions of hours
1.0040 **	13.57	0.9988	0.52	annual effective working time in hours
1.0050 **	14.20	0.9983	0.83	annual working time per PCGA in hours

*) to avoid multicollinearities, labour market indicators were subsequently included into the baseline model

variable changes by one unit. The column 'diff Chi2' indicates the reduction of the Chi²-value when the variable is excluded from the model. The decrease of the Chi²-value is therefore a measure of the relative influence of a parameter.

4.1 The Model of Return to Employment

First of all, the results will be reported for the baseline model for the return to employment. Here, men have a 1.7-fold higher rate than women. The coefficient verifies well known inequities towards women in working life. Surprisingly, we find a negative relation between level of education and transition rate. The rate for persons with degrees which entitle entrance to a university or senior technical college is only about 60% of the rate for the remaining population. The reasons for this finding must be speculative. Employees with higher degrees may have problems finding an appropriate entry-level. A correlation between education and sex may also be responsible. The result, however, corresponds with the occupational status coefficients for blue-collar workers who have on the average relatively lower educational degrees than persons from other occupational states. Blue-collar workers show a 1.3-fold higher rate than the rest. The age effect ($\alpha=0.7067$) for persons older than fifty confirms the reemployment problems of older unemployed persons. The variable age had to be included into the model as a set of dummy variables because exploratory analysis showed the relationship between rate and age to be non-linear in the logarithms. Age less than 50 does not include significantly different sub-groups. Women who had had a child within the last 12 months before unemployment show a significant lower rate than the remaining population. This may be due to true reemployment problems. For example, young mothers who want to return to employment may be primarily interested in part-time jobs which are relatively rare. Job

search theory also suggests that material and, more importantly, immaterial search benefits to young mothers are higher than for other persons, implying lagged resumption of employment. Material search benefits arise for instance from saved expenditures for child care. Immaterial search benefits result mainly from the sociological fact that the growing occupational activity of women has not lead to an equivalent adaptation of traditional sex roles. Therefore, full-time jobs mean a considerable burden for young mothers, as long as there does not exist an extended family which may act to assist. The state of unemployment may exonerate women from such problems. The two above-mentioned causes for a reduced rate of young mothers are consistent, since the interest in part-time jobs may be viewed as an optimum search plan.

Particular interest should be paid to the effect of Arbeitslosengeld,*) which indicates that persons with an existing claim have a clearly increased rate compared to other persons or periods without a claim. Initially, this finding appears counter-intuitive. According to the arguments of job search theory it probably would have been expected that people without any benefit claims would be most urgently forced to an early termination of unemployment. This is not the case in our finding. According to human capital theory the present result could be explained by the circumstance that persons entitled to benefit claims have more vocational experience at their disposal than others due to preceeding occupation. In contrast, the majority of the population of unentitled persons consists of entry-level employees or people with a long interruption of their labour force participation. In any case, the coefficient for Arbeitslosengeld does not yield evidence of an increased propensity to remain unemployed with higher levels of benefit claims, assuming that an interest in re-occupation exists. This result is

---

*) see Figure 1 in Chapter 2

supported by the fact that the coefficient for the level of social credit points did not turn out to be significant, although the number of social credit points can only be taken as a rough indicator for the actual level of unemployment compensation as was mentioned above. Finally, we controlled the effect of compensation by a dummy variable, which was set to unity, when the end of the time limit for Arbeitslosengeld had been reached, and to zero otherwise.*) The corresponding coefficient did not prove to be significant, which supports the conclusion above: there is no extra increase in the rate when benefit claims expire.

A look at the coefficients for the labour market indicators (Table 4) shows the expected results. Increasing tightness of the labour market correlates positively with unemployment duration. For example, an increase in the annual unemployment rate by 1% leads to a decrease of the estimated rate by a factor of 0.88, which corresponds to an extension of mean unemployment duration by 14%. The number of employees, excluding civil servants, as well as the number of economically active persons do not fit the model variables very well. Both indicators show a decrease after 1973 and an increase after 1977, while unemployment duration increased continously over the whole observation period. It is obvious that the observable increase of employment indicates a relaxation of the labour market which had been counterbalanced at the same time by a disproportionate increase of the population capable of gainful activity or a decrease of the total number of man hours worked.

4.2 The Model of Withdrawal to Non-activity

Tables 3 and 4 illustrate that the process of returning to employment is principally different from the process that leads to retirement from working life. In contrast to the

---

*) see Figure 1 in Chapter 2

preceeding case, the dominant part in the transition to non-activity is played by institutional regulations concerning claims for unemployment compensation. There are several reasons for this interpretation. We find that entitled persons have a relatively low rate as long as a claim for Arbeitslosengeld exists. Upon reaching the time limit of their claim for Arbeitslosengeld the transition rate to non-activity for this population rises to a level more than four times higher compared to other periods. This result is understandable if one realizes that for retiring persons who are interested in windfall profits from existing benefit claims, a maintenance of unemployment registration beyond the time limit of entitlement would only be a useless effort. Beyond this, both effects produce the highest Chi^2-values within the submodel. Alone the exclusion of the parameter for reaching the time limit of entitlement would reduce the Chi^2-value of the submodel by half.

In addition, a strong argument from our view is that not a single labour market indicator has a significant influence within the model. From this it may be concluded that the observed retirement behaviour remained constant over the observation period and occured independently from the labour market.

Nearly 70% of the transitions that occur at the end of the entitlement period to Arbeitslosengeld lead to non-activity, while the overall ratio between transitions to non-activity and transitions to employment is about 3:5. The proportion of women among retiring persons at that point in time reaches almost 90%, while there is no sex effect among persons who return to employment at the end of their eligibility for unemployment insurance. This shows that a resumption of employment at the end of entitlement to Arbeitslosengeld is difficult to control by individuals themselves.

It should be mentioned that the proportion of young mothers with benefit claims who retire amounts to about 56%. From this we conclude that motherhood alone is a strong, but not necessarily the only reason to leave the labour market. However, among young mothers who are entitled to Arbeitslosengeld and leave unemployment when reaching the time limit for Arbeitslosengeld, the number of transitions to non-activity exceeds the number of transitions to occupation by almost seven times.

For the above-mentioned reasons it is understandable that within our model of transition to non-activity there are no significant effects of sex and recent motherhood, due to their correlatedness with the effects of unemployment compensation. The interpretation of the occupational status coefficients should be as follows: a reduced rate for blue-collar and white-collar workers compared to the remaining population, will be caused by the fact that the reference group mainly consists of self-employed and other persons not entitled to benefit claims. The effect has to be seen in relation to the fact that persons with benefit claims show a clearly increased rate when the time limit for Arbeitslosengeld is reached.

Without further evidence we cannot ascertain whether the coefficients for educational degrees may be interpreted in the same way. Nevertheless, it seems plausible that among the subpopulation of retiring persons, more than two-thirds of which are women, those women with higher educational degrees persist a relatively long time in unemployment. Moreover, it is assumed that these women are really interested in vocational reintegration. Possible resignation leading to retirement may occur relatively late in these cases and be independent of the existence of entitlement to unemployment compensation. Such connections, however, may only be supposed here, a verification would require far more detailed data.

The remaining age effect also cannot be uniquely interpreted. It has already been mentioned that the phenomenon of a prematurely drawn pension claim, via unemployment compensation, could not be observed here due to restrictions of the sample design. It may nevertheless be assumed that persons potentially falling into such a pattern are present in our data as right-censored cases. So far, it remains open as to whether the reduction in the rate by about 50% for persons older than fifty is produced by this pattern or whether we have here anew an interaction with effects related to benefit claims. Strictly speaking the two views are not mutually exclusive.

5. Summary and Final Remarks

Our analysis shows that the transition from unemployment to employment differs fundamentally from the transition to non-activity after unemployment. This is expressed in the result that the transition to non-activity is primarily directed by institutional regulations of unemployment compensation, while the process of return to employment is dominated by other factors. When reaching the time limit of their claim for Arbeitslosengeld the transition rate to non-activity for persons entitled to benefit claims exceeds by four times the rate during other periods. This is supported by the fact that labour market indicators do not have any influence on the transition to non-activity while on the other hand the transition rate of unemployment to employment decreases with increased tightness on the labour market. Bottlenecks in the labour market may prevent persons who are searching for a job from getting one. This is not the case for persons who decide to leave the labour market. A possible interest in windfall profits from unemployment insurance among the latter is independent of economic trends and only related to the time limit of unemployment benefit claims.

Distinguishing retirement from re-employment throws new light on conclusions from job search theory. It is often proposed to use the level of unemployment compensation as a regulatory tool to cure unemployment. Our results, however, cast doubts on the effectiveness of such measures. There is no evidence that a lowering of unemployment compensation would help to alleviate the problems of return to employment, although it would surely increase the problems of the unemployed. Even people who leave for non-activity and are interested in windfall profits may not be expected to respond to the level of compensation and unemployment duration will be unattached, since the decisive effect seems to be the time limit of compensation.

The estimates for the return to employment are impaired to some extent by the fact that the chosen labour market indicators do not produce the relatively strongest effects within the submodel. From the relative weight of the coefficients it might be concluded that unemployment can primarily be viewed as an individual problem. In our opinion, however, this consequence seems to be too premature. It is not clear yet, for example, to which extent models of the present form include the risk of committing an ecological fallacy, a common problem as are associated with aggregate data (Küchler 1979). Unfortunately, the existing data set gives few opportunities for a more differentiated analysis of the aggregate variables, either by regionalization or by disaggregation.

It seems remarkable, in this context, that estimations using the labour market indicators calculated instead on a monthly basis did not lead to significant coefficients. Seasonal fluctuations of the labour market indicators apparently do not have the same effect as long-term changes. The range of seasonal fluctuations within a single year often surpasses the range of mean changes between subsequent years. As a result, one of the remaining tasks will be to produce better

fitting indicators for structural problems on the labour market.

Literature:

Andreß H.-J. 1985: Multivariate Analyse von Verlaufsdaten. Mannheim

Arminger G. 1984: Analysis of Event Histories with Generalized Linear Models. In: Diekmann A., Mitter P. (Eds.): Stochastic Modelling of Social Processes. Orlando, pp. 245-282

Bailey D., Parikh A. 1985: An Analysis of Duration of Unemployment: 1967 - 1979. In: Empirical Economics, Vol. 10, pp. 131-142

Büchtemann C.F. 1985: Recurrent Unemployment - Longitudinal Evidence for the Federal Republic of Germany. SAMF-Arbeitspapier Nr. 1985-3, Berlin

Büchtemann C.F., Infratest Sozialforschung 1983: Die Bewältigung von Arbeitslosigkeit im zeitlichen Verlauf. Bonn

Buttler G. 1984: Arbeitsmarktanalyse, Aufgaben und Möglichkeiten. In: Buttler G. (Ed.): Arbeitsmarktanalyse. Göttingen, pp. 7-16

Cox D.R., Oakes D. 1984: Analysis of Survival Data. London-New York

Diekmann A., Mitter P. 1984: Methoden zur Analyse von Zeitverläufen. Stuttgart

Egle F. 1984: Arbeitsmarktindikatoren. In: Buttler G. (Ed.): Arbeitsmarktanalyse. Göttingen, pp. 53-67

Feldstein M.S. 1975: The Importance of Temporary Lay-Offs: An Empirical Analysis. In: Brookings Papers of Economical Activities, pp. 725-744

Freiburghaus D. 1978: Dynamik der Arbeitslosigkeit. Meisenheim am Glan

Friedrich W., Ronning G. 1985: Technischer Fortschritt - Auswirkungen auf Wirtschaft und Arbeitsmarkt. In ifo-Schnelldienst, 22/85, pp. 13-25

Hamermesh D. 1977: Jobless Pay and the Economy. Baltimore

Hausman J. 1978: Specification Tests in Econometrics. In: Econometrica, Vol. 46, pp. 1251-1271

Helberger C.: 1982: Arbeitslosigkeit als Gegenstand mikro-ökonomischer Theorien zur Funktionsweise der Arbeits-märkte. In: DIW-Vierteljahreshefte zur Wirtschafts-forschung, Heft 4/82, pp. 398-412

Hujer R., Schneider H. 1986: Ökonometrische Ansätze zur Analyse von Paneldaten: Schätzung und Vergleich von Übergangsratenmodellen. Sfb 3-Arbeitspapier Nr. 190, Frankfurt-Mannheim

Kalbfleisch J.D., Prentice R.L. 1980: The Statistical Analysis of Failure Time Data. New York

Kiel W. 1985: Datenhandbuch zur Lebenslagen-Studie

Küchler M. 1979: Multivariate Analyseverfahren. Stuttgart

Lawless J.F. 1982: Statistical Models and Methods for Lifetime Data. New York

Lempert-Helm I. 1985: Zugangs- und Verbleibsrisiken der Arbeitslosigkeit im Zeitvergleich. In: Hujer R., Knepel H. (Eds.): Mobilitätsprozesse auf dem Arbeits-markt. Frankfurt-New York, pp. 193-222

Lippman S.A., McCall J.J 1976: Job Search in a Dynamic Economy. In: Journal of Economic Theory, Vol. 12, pp. 365-390

Moffitt R. 1985: Unemployment Insurance and the Distribution of Unemployment Spells. In: Journal of Econometrics, Vol. 28, pp. 85-101

Tuma N.B., Hannan M.T. 1984: Social Dynamics. Orlando

II Family formation, migration and fertility

Gender differences in family formation: Modelling the life course specificity of social differentiation

Georgios Papastefanou

0. Problemstellung

Wenn man nach Variationen des Zeitpunktes der Famliengründung sucht, dann fällt vor allem der Unterschied zwischen Männern und Frauen auf. Frauen heiraten früher als Männer. Dies gilt - wenn man das mittlere Lebensalter bei der ersten Heirat von Männern und Frauen vergleicht - 1982 genauso wie 32 Jahre zuvor (siehe Schaubild 1). In der Zeit von 1950 bis 1975 sank das durchschnittliche Heiratsalter, der mittlere Altersunterschied der beiden Populationen blieb dabei erstaunlich konstant. Dies kontrastiert mit den sozialen und wirtschaftlichen Veränderungen, die in der Bundesrepublik Deutschland nach dem Zweiten Weltkrieg das Verhältnis der Geschlechter betroffen haben. Insbesondere bei Frauen hat sich der relative Stellenwert von beruflichen und familialen Ereignissen im Lebensverlauf im Zuge der Herausbildung und Modernisierung der westdeutschen fortgeschrittenen Industriegesellschaft gewandelt. Durch die sogenannte "Bildungsexpansion" wurde in dieser Periode das Bildungs- und Qualifikationsprofil von Frauen dem der Männer angeglichen.

Da mit dem Berufszugang einerseits die Chancen und die Anforderungen des weiteren Berufweges erhöht wurden, andererseits die dort getroffenen Entscheidungen Konsequenznen für Stellenwert und Zeitpunkt der Familienbildung haben, stellt sich die Frage, in welchem Maße die vorgezogene Entscheidung zur Familie bei

Frauen Ausdruck ihrer im Vergleich zu Männern schlechteren Bildungschancen ist. Offenbar gibt es nicht einen einfachen Zusammenhang zwischen Berufszugang und Familiengründung, da trotz der erweiterten Bildungsbeteiligung das Heiratsalter von Männern und Frauen gesunken ist.

Die vorliegende Arbeit untersucht, in welchem Verhältnis der allgemeine und berufliche Bildungsweg von Frauen und Männern zu einer "kritischen" Heiratsphase steht und wie sich dieses Verhältnis bei ausgewählten Kohorten gewandelt hat. Sie bedient sich hierbei der stochastischen Modellierung der Wartezeit bis zur ersten Heirat auf der Basis der log-logistischen Verteilung.

1. Konzeptuelle Analyse

1.1 Heiratsentscheidung und Familiengründung

"Familie" als soziales Phänomen hat in der westlichen Industriegesellschaft der Neuzeit eine zweifache Bedeutung (vgl. LÜSCHER 1985): a) mit Familie ist eine lebensgemeinschaftliche Gruppe gemeint, die in der Regel einen gemeinsamen Haushalt führt. In diesem Sinne entsteht eine Familie dann, wenn ein Haushalt um eines oder mehrere Kinder erweitert wird. Unter dieser Perspektive ist deshalb die Geburt des ersten Kindes jenes Ereignis, das zur Bildung einer Familie führt. b) Innerhalb des Institutionengefüges der westlichen Gesellschaften ist die Familie eine institutionalisierte Beziehungsform, deren rechtlich-sozial notwendiges Definitionsmerkmal die formale Eheschließung nach dem bürgerlichen Gesetzbuch ist. Eine Familiengründung wird in diesem Rahmen durch die Eheschließung vollzogen, auch wenn kein Kind gezeugt wird. (Auf das Verhältnis zwischen rechtlich-staatlicher Normensetzung und lebensweltlichem Sinngehalt der Definitionen der Familiengründung soll hier nicht eingegangen werden).

In diesem Sinne ergeben sich zwei analytisch zu trennende Ereignisse, die in der Familiengründung eine gemeinsame Handlungsbedeutung haben. Durch die beiden Ereignisse werden unterschiedliche soziale Zusammenhänge indiziert. Während der

erste Aspekt ein universelles Merkmal von Familie darstellt, ist der zweite Aspekt ein genuin sozial geformtes Merkmal von Familie. Dies bedeutet, daß bei Analysen mit historischen Bezügen Veränderungen der sozialen Bedeutung der Eheschließung mitberücksichtigt werden müssen. Man kann die Eheschließung einerseits als einen mit rechtlichen und sozialen Sanktionen verbundenen Abschlußritus einer heterosexuellen Beziehung betrachten. Die Eheschließung als sozialer Vertrag schließt den Prozess der Partnerwahl ab, indem die informelle, nur psychisch-soziale Beziehung zur rechtlichen Bindung in einem gemeinsamen Haushalt transformiert wird. Andererseits ist die Eheschließung ein Initiationsritus, der von der Zweierbeziehung zur Familie überleitet, indem auf der Basis der formalisierten Bindung des Paares die gegenseitigen Ansprüche und Pflichten von Mann, Frau und möglichen Nachkommen legitimiert werden. Die Zeugung eines oder mehrerer Kinder ist in hohem Maße an die Eheschließung gekoppelt. Erst wenn die Beziehung des Paares legalisiert wird, erhält auch die Beziehung von Mann und Kindern eine rechtliche und darum kalkulierbare und sozial stabile Dimension. Dieser Tatbestand spiegelt sich darin wider, daß die Zeugung eines Kindes in der westdeutschen Gesellschaft normativ nach wie vor an die feste Bindung des Paares geknüpft ist. D.h. daß die Entscheidung zur Zeugung eines Kindes auch die Entscheidung zur festen Bindung und Haushaltsgründung impliziert. Diese beiden Merkmale sind bis zu einem gewissen Grad normativ miteinander verbunden und machen die Realität einer Familiengründung aus. Allerdings ist das Bedingungsverhältnis von Heirat bzw. Haushaltsgründung und Zeugung des ersten Kindes ein sozio-ökonomisch beeinflußtes Verhältnis. Die neueste Entwicklung der Familie ist gerade davon gekennzeichnet, daß die Entscheidung zu einem gemeinsamen Haushalt in immer höherem Maße unabhängig von der Entscheidung zur Familiengründung wird (TYRELL 1985, NAVE-HERZ 1984).

Im Rahmen der vorliegenden Arbeit gehen wir davon aus, daß Heiratsentscheidung und Entscheidung zur Zeugung des ersten Kindes motivational sehr eng assoziiert sind. Familiengründung ist das Ergebnis einer Paarbeziehung, die unter der Norm der Neolokalität zur Eheschließung bzw. gemeinsamen Haushaltsführung

und Zeugung von Kindern führt. Sie umfaßt folgende Aspekte:
1) psycho-sexuelle Bindung, 2) gemeinsamer Haushalt, in dem die
Befriedigung ökonomischer, psychisch-sozialer und sexueller
Bedürfnisse auf Dauer sichergestellt wird, 3) Eheschließung als
rechtlich-soziale Stabilisierung und Rationalisierung der
psycho-sexuellen Beziehung, 4) Zeugung von Kindern.

1.2 Eheschließung: Entscheidung, Ereignis und Prozeß

Die Eheschließung ist das Ergebnis eines komplexen, sozial
eingebetteten individuellen Entscheidungsprozesses. Man kann
fragen, wovon die Entscheidung zur Heirat abhängig ist. Dabei
kommt man in der Regel zu dem Ergebnis, daß die Eheschließung
zumindest in der Bundesrepublik Deutschland ein weit verbreitetes
Lebensmuster darstellt. Nahezu jede Frau und jeder Mann heiraten
und gründen irgendwann im Laufe ihres Lebens eine Familie.

Unter unserer Perspektive stellt diese Betrachtungsweise eine
statische Betrachtung dar. Davon abzugrenzen ist eine dynamische
Perspektive, deren paradigmatischer Ausgangspunkt die Frage nach
dem Zeitpunkt einer Handlung ist. Unter dieser Sichtweise stellt
sich das Verständnis von der individuellen Handlungsentscheidung
zur Eheschließung in anderer Weise dar.

In einer Verlaufsperspektive kann man die Summe der Handlungen,
die die Eheschließung ausmachen, als Ereignis bezeichnen, das den
einen Zustand beendet und einen neuen Zustand initiiert. Wenn wir
uns hier mit dem Heiratsprozeß beschäftigen, so bedeutet dies,
daß wir vor allem jenen Zustand untersuchen, den das
Heiratsereignis abschließt. Die qualitative Beschreibung dieses
Zustandes, das " Ledig Sein", müßte sehr differenziert erfolgen,
da dieser Handlungszustand bis zur Heirat nicht immer gleich
bleibt. Indem er als Zustand des Nicht-Verheirateten mit
entsprechenden Handlungschancen und -zwängen bezeichnet wird,
reicht er theoretisch von der Geburt bis zum Ereignis. In
westlichen Gesellschaften ist es aber trivial, daß das Risiko
einer Heirat nicht ab Geburt, sondern erst ab einem sozial
bestimmten Zeitpunkt vorhanden ist. Mit anderen Worten: Das
Risiko einer Heirat beginnt mit dem Zeitpunkt, zu dem ein für die

Eheschließung sozial relevantes Ereignis eingetreten ist. Von
diesem Moment an gibt es für den individuellen Akteur eine
bestimmte Wahrscheinlichkeit zu heiraten, wobei die
Heiratswahrscheinlichkeit durch einen bestimmten zeitlichen
Verlauf gekennzeichnet ist. Dies soll im folgenden als
"individueller Heiratsprozeß" bezeichnet werden. Es ist zu
beachten, daß man sich die individuelle Wahrscheinlichkeit immer
innerhalb der spezifischen Population, zu der der Akteur
angehört, vorzustellen hat. In diesem Sinne ist der Heiratsprozeß
ein Person-Populationsprozeß (vgl. FEATHERMAN/LERNER 1985).

Wenn man also eine dynamische Perspektive an die
Heiratsentscheidung legt, findet ein Übergang von einem
deterministischen zu einem probabilistischen Ansatz statt, dessen
zentraler Begriff der Heiratsprozeß ist. Die Frage nach der
geschlechtsspezifischen Heiratsentscheidung, die nahezu ohne
Aussagewert ist, weil letztlich fast alle Männer und Frauen
heiraten, wird zur Frage nach der Geschlechterdifferenzierung des
Heiratsprozesses. Allerdings benötigt man bei der Untersuchung
eines Prozesses immer auch eine Angabe über den Beginn des
Prozesses. Dies heißt im Hinblick auf den Heiratsprozeß, daß es
notwendig ist, den Anfangszeitpunkt des initiierenden Ereignisses
theoretisch zu spezifizieren, um den Heiratsprozeß zu beschreiben
und Faktoren seiner Variabilität zu untersuchen. Die
Spezifikation des Heiratsprozesses benötigt also - neben dem
Zielereignis - auch die Beschreibung des initiierenden
Ereignisses.

In einer sozialwissenschaftlichen Perspektive bieten sich
mehrere Ereignisse an, die als Beginn eines Heiratsprozesses
verstanden werden können. 1) sexuelle Reife: Pubertät; 2) Ende
der Adoleszenz; 3) Eintritt in Heiratsmündigkeit; 4) Austritt aus
der Schule; 5) Ende der beruflichen Ausbildung; 6) Beginn der
Erwerbstätigkeit.

Diese Liste erhebt nicht den Anspruch auf Vollständigkeit oder
darauf, den Beginn des Heiratsprozesses theoretisch zu
durchdringen. Durch Angabe dieser Ereignisse, die z.T.
operationale und z.T. theoretische Ereignisse darstellen, soll

klar werden, daß je nach initiierendem Ereignis unterschiedlich
lange Heiratsprozesse spezifiziert werden. Je nach dem, ob man
die Pubertät als Beginn des Heiratsrisikos oder den Beginn der
ökonomischen Selbständigkeit ansetzt, haben wir es mit einem
unterschiedlich heterogenen Heiratsprozeß zu tun. In dem einen
Fall sind vor allem Partnerwahlprozesse in den Heiratsprozeß
involviert, in dem anderen Fall sind diese eher vom Heiratsprozeß
abgelöst. Ersteres beschreibt den Partnerwahlprozeß als einen
Subprozeß des Heiratsprozesses, bei letzterem handelt es sich um
einen Heiratsprozeß, der im Anschluß an die erfolgte Partnerwahl
einsetzt. Wie lassen sich diese Begriffe empirisch einlösen?

Als Beginn des Heiratsprozesses setzen wir in dieser Arbeit den
Zeitpunkt der Heiratsmündigkeit. Damit berücksichtigen wir von
vornherein die Geschlechtsunterschiede, die durch diese rechtlich
fixierte Norm entstehen. Das Alter der Heiratsmündigkeit beträgt
für Frauen 16 Jahre und für Männer 18 Jahre. Die bedingte
Heiratswahrscheinlichkeit, d.h. die Relation von Personen, die de
facto heiraten, zu Personen, die potentiell heiraten könnten,
stellt eine adäquate Operationalisierung der individuellen
Heiratschance oder -bereitschaft dar. Dieser Variablen entspricht
die Heiratsrate, wie sie in den Statistischen Jahrbüchern
mitgeteilt wird. Der Heiratsprozeß ist für das Individuum mit der
Eheschließung beendet. Als Populationsprozeß ist er entweder dann
beendet, wenn alle dazugehörenden Personen geheiratet haben, oder
zu dem Zeitpunkt, ab dem die Heiratswahrscheinlichkeit sich nicht
mehr verändert.

Am Beispiel des Jahrganges 1940 wollen wir betrachten, in
welcher Weise sich die individuelle Heiratsbereitschaft bzw.
-chance von Männern und Frauen in ihrem Lebensverlauf
unterscheidet (siehe Schaubild 2). Wir wollen zwei Merkmale des
Heiratsprozesses dieses Jahrganges hervorheben, die von
spezifischer Bedeutung für die weitere Argumentation sind: 1. Die
Wahrscheinlichkeit einer Eheschließung ist nach der
Heiratsmündigkeit nicht über das Lebensalter gleichmäßig, sondern
nicht-monoton verteilt. Dieses Verteilungsmuster findet man
sowohl bei Männern wie auch bei Frauen. 2. Frauen haben in einer
zeitlich begrenzten Phase des Prozesses eine höhere

Heiratswahrscheinlichkeit als Männer.

Man sieht deutlich, daß bei Männern der Heiratsprozeß mit dem 18. Lebensjahr einsetzt; bei den Frauen dagegen befindet sich der Heiratsprozeß in diesem Alter schon ein Stück weit in der "Beschleunigungsphase" . Bis zum 27. Lebensjahr haben Frauen eine höhere Heiratswahrscheinlichkeit als Männer, aber danach ist der Unterschied so klein und variabel, daß man ihn vernachlässigen kann. Für Frauen ist in jüngeren Lebensjahren die Heiratswahrscheinlichkeit höher als für Männer. Ab dem 28. Lebensjahr haben Männer und Frauen die gleichen Chancen zu heiraten. Bei dieser Darstellung wird deutlich, daß auch die Geschlechtsspezifizität der Heiratschance als Phänomen selbst altersabhängig ist. Es liegt eine Interaktion der Effekte von "Lebensalter" und "Geschlechtszugehörigkeit" in bezug auf die Heiratswahrscheinlichkeit vor.

1.3 Überlegungen zur Altersabhängigkeit der Familiengründung

1.3.1 Biologische und biosoziale Zusammenhänge

Wie kann die glockenförmige Verteilung der Heiratswahrscheinlichkeit unter Rückgriff biologischer Zusammenhänge erklärt werden? Halten wir zunächst noch einmal fest, mit welchen Fakten wir es hier zu tun haben. a) In einer Population verläuft die körperliche Reifung nicht homogen, sondern weist eine spezielle Altersverteilung auf. b) Mädchen erreichen den Zustand der körperliche Reife früher als Jungen. c) Die körperliche Reife impliziert bei Mädchen zwei Aspekte: Sexualität und Fruchtbarkeit; bei Knaben handelt es sich bei der Reifung um die Entwicklung sexueller Bedürfnisse d) Mit der körperlichen Reife geht bei Mädchen die Entwicklung von reproduktionsnahen Fähigkeiten einher.

Wenn wir uns unter diesen Gesichtspunkten das Schaubild 2 anschauen, dann können wir den ersten Teil der Wahrscheinlichkeitsverteilung so interpretieren, daß er die Verteilungsform der körperlichen Reifung abbildet. Der

vorverlagerte Heiratsprozeß bei Frauen entspricht ihrer früheren sexuellen Entwicklung, die sie zugleich zur Mutterschaft befähigt. Folgende Aspekte der vorliegenden Verteilung lassen sich jedoch nicht im Rahmen eines rein biologischen Ansatzes erklären: a) Sowohl bei Männern wie auch bei Frauen sinkt die Wahrscheinlichkeit der Heirat nach Erreichen einer bestimmten Altersstufe wieder ab. b) Der Heiratsprozeß der Frauen ist in stärkerem Maße auf einen bestimmten Altersbereich konzentriert als der der Männer. Die Familiengründung der Männer ist weniger stark auf ein bestimmtes Lebensalter festgelegt. Unter Einbeziehung der evolutionären Perspektive werden diese Aspekte der Altersverteilung der Heiratswahrscheinlichkeit interpretierbar. Demnach ist die sexuelle Aufmerksamkeit der Männer primär auf Frauen jüngerer Altersgruppen ausgerichtet. D.h. aber auch, daß Frauen in dem Abschnitt ihres Lebensverlaufs die höchsten Heiratschancen haben, in dem sich ihre Reproduktionsfähigkeit entfaltet. (DAVIS/vanden OEVER, 1982:506). Das hat zur Folge, daß die Wahrscheinlichkeit von Frauen, vor oder nach diesem Altersabschnitt zu heiraten, eher geringer sein wird. Auf diese Weise entsteht bei Frauen ein glockenförmiger Heiratsprozeß wie im Schaubild 2. In diesem Sinne spiegelt die Form ihres Heiratsprozesses ihre sexuelle Attraktivität für Männer wider, und man könnte die lebensaltersspezifischen Heiratswahrscheinlichkeiten als die sexuellen Marktwerte von Frauen interpretieren (COLEMAN 1984).

Die körperliche Reifung bedeutet für Frauen auch die Entwicklung von Schwangerschaftsmöglichkeiten und -fähigkeiten, die nach einem bestimmten Lebensalter wieder reduziert werden. Unter der Annahme, daß Frauen in der Regel ein Kind wünschen, ist bei ihnen der Familiengründungsprozeß relativ altershomogen, weil für alle Frauen die Möglichkeit einer Schwangerschaft auf eine bestimmte Lebensspanne beschränkt ist. Wenn aber die Sexualität von Männern relativ altersunabhängig ist (nach der Reifung), so müßte dies zu einer im Alter beliebigen Chance der Heirat und Familiengründung führen.

Wenn man den Heiratsprozeß der Männer mit dem der Frauen vergleicht, bestätigt sich diese Annahme in etwa.

Die Altersunabhängigkeit der Heiratswahrscheinlichkeit von
Männern ist jedoch nicht so deutlich ausgeprägt, als daß ein
altersbeliebiges, sexuelles Interesse der Männer konsistent eine
umfassende Erklärung auch ihres eigenen Heiratsprozesses bieten
könnte.

Zusammenfassend läßt sich festhalten: Der physiologische
Reifungsprozeß von Sexualität und Fekundabilität hat zur Folge,
daß sich die Wahrscheinlichkeit der Familiengründung über den
Lebensverlauf glockenförmig verteilt. Aus zwei Gründen betreffen
die unterschiedlichen Heiratschancen von Männern und Frauen vor
allem die Reifungsphase: 1. Männer zeichnen sich durch eine
monosexuelle Orientierung aus. 2. Bei Frauen sind zwei Aspekte
relevant: die körperliche Reifung setzt schneller ein, und die
Fekundabilität ist auf einen bestimmten Altersbereich beschränkt.
Aus diesen Gründen unterscheiden sich die
Heiratswahrscheinlichkeiten von Männern und Frauen hauptsächlich
in der Reifungsphase. Unbefriedigend bleibt innerhalb dieses
Ansatzes die Erklärung für die ab einem bestimmten Alter sinkende
Heiratswahrscheinlichkeit bei Männern.

1.3.2 Sozio-kulturelle Erklärung

Der Kern dieses Erklärungsansatzes besteht in der Annahme, daß
es eine Altersnorm der Familiengründung (NEUGARTEN/DATAN 1973,
HAGESTAD/NEUGARTEN 1985, MODELL u.a. 1978) gibt. Durch die
Erwartungshaltung von Eltern als auch von peers wird ein
bestimmtes Lebensalter als der angemessenste bzw. als der
späteste Zeitpunkt der Familiengründung definiert. Eine solche
soziale Norm ist dadurch charakterisiert, daß zum einen ein
spezifischer Altersbereich spezifiziert, zum anderen ihr eine
gewisse handlungsleitende Stärke beigemessen wird. Die Bedeutung
dieser Altersnorm wird darin deutlich, daß in einem bestimmten
Lebensalter Heiratswahrscheinlichkeit maximal ist. In diesem
Sinne reflektiert der Unterschied zwischen den Heiratsprozessen
beider Geschlechter die Geltung unterschiedlicher Altersnormen.
Es bleibt die Frage offen, ob es für Männer eine andere
Altersnorm gibt als für Frauen oder ob es nur für Frauen eine
konkret erfahrbare Altersnorm gibt.

Wenn man Schaubild 2 auf dem Hintergrund der Annahme
interpretiert, daß hier die Wirkung einer Altersnorm zum Ausdruck
kommt, dann müßte man sagen, daß für Frauen die Norm eines frühen
Heiratsalters mit größerer Sanktionskraft verbunden ist als für
Männer. Die Heiratschance von Männern unterscheidet sich von
jener der Frauen nur in der spezifischen Lebensphase, in der die
differentielle normative Sanktionierung der Heiratsentscheidung
ihren Höhepunkt hat. In dem Moment, von dem an das soziale
"Kräftefeld" der Altersnorm qua Lebensalter überwunden ist,
gleichen sich Heiratsbereitschaft bzw. Heiratschancen von Männern
und Frauen an. Die handlungsleitende Wirkung einer solchen
Alternorm basiert im wesentlichen auf der primären Sozialisation.
Die geschlechtsspezifische Sozialisation ist dadurch
gekennzeichnet, daß die Orientierung von Mädchen auf familiäre
Aktivitäten ausgerichtet wird, die der Männer auf Aktivitäten in
beruflichem Bereich. Die Tatsache, daß der Heiratsprozeß von
Frauen in einem früheren Lebensalter verläuft und
altershomogener ist, ließe sich dadurch erklären, daß in der
Kindheit von Frauen das Bedürfnis nach Geselligkeit und Nähe,
sowie ihre Ausrichtung auf Konformität und Kollektivität
besonders stark sanktioniert wurden.

Das stärkere Bedürfnis nach Familie und die größere Unterordnung
der persönlichen Entwicklung unter die Erfordernisse der
Familiengründung - beides führt bei Frauen sowohl zum schnelleren
wie auch zum alterskonformeren Eintritt in die Familie, als dies
bei Männern der Fall ist. Bei den Männern dagegen liegt der
Schwerpunkt auf Ausbildung von Selbstständigkeit, die zu einem
späten Heiratsprozeß mit einer relativ großen Variation führt.

Damit ist die Frage noch nicht beantwortet, warum ab einem
bestimmten Alter die Heiratsbereitschaft von Männern und Frauen
auf das gleiche niedrige Niveau absinkt. Im Zusammenhang mit dem
Norm-Konzept sind zwei Antworten möglich: 1. Ab einem bestimmten
Alter läßt die Wirkung der sozialen Kontrolle nach und verliert
ihren zwischen Männern und Frauen differenzierenden Effekt. Da
Frauen in höherem Maße der sozialen Kontrolle ausgesetzt sind,
ist dieser Effekt wirkungsvoller als bei Männern. Dies führt

dazu, daß die Heiratsbereitschaft von Frauen und Männern sich
nicht mehr unterscheidet. 2. Die sozialisatorische Prägung wirkt
sich nicht über die gesamte Lebensspanne aus, sondern wird mit
fortschreitender Biographie überlagert und verändert durch die
Akkumulation neuer Erfahrungen. Dies reicht jedoch nicht aus, um
die Angleichung der Heiratsbereitschaft befriedigend zu erklären;
denn warum sollte bei Frauen im gleichen späten Lebensalter wie
bei Männern die in der Sozialisation erworbenen Dispositionen
stärker und rascher überwunden werden als bei den Männern?

1.3.3 Sozialstrukturelle Erklärung

Die sozialstrukturelle Erklärung basiert im wesentlichen auf der
statischen Annahme, daß es eine festgelegte konsistente familiale
Arbeitsteilung von Männern und Frauen gibt. In der Familie
übernehmen Frauen die expressiven und reproduktionsnahen Rollen
und Männer die Rolle des Versorgers. Der Beitrag, den Männer zur
Familiengründung zu leisten haben, besteht also darin, daß sie
die ökonomische Sicherheit der Familie gewährleisten. Daraus
liesse sich die Annahme ableiten, daß Männer erst später als
Frauen eine Familie gründen können, da sie zuerst die
instrumentelle Aufgabe der Existenzsicherung der zukünftigen
Familie lösen müssen. Eine genuine Altersabhängigkeit der
Heiratswahrscheinlichkeit gibt es nach dieser Vorstellung nicht.
Die Altersvariation der Heiratschance ist bloßer Ausdruck der
sozio-ökonomischen Heterogenität bzw. der sozio-ökonomischen
Ungleichheit der Männerpopulation. Die insgesamt höhere
Heiratschance von Frauen kann ebenfalls als Ausdruck
gesellschaftlicher Ungleichheit verstanden werden, da Frauen
insgesamt weniger Anteil an instrumentellen Aktivitäten haben.

Wie aber kann man mit diesem Ansatz die Art und Weise erklärt
werden, wie die Heiratschancen der Frauen sich in ihrem
Lebensverlauf verteilen? Wenn man eine statische Auffassung
zugrunde legt, läßt sich diese Frage leicht beantworten: Frauen
haben größere Heiratschancen als Männer, weil sie weniger
instrumentelle Aufgaben übernehmen. Damit ließe sich sowohl die
Tatsache der höheren Heiratswahrscheinlichkeit (siehe Schaubild
2) erklären als auch, daß der Heiratsprozeß bei Frauen früher
einsetzt als bei Männern.

Daran knüpft die Frage an, ob auch die Altersvariation der Heiratschancen von Frauen auf diese Weise erklärt werden kann? Wenn man davon ausgeht, daß auch bei Frauen die Fähigkeit zur Existenzsicherung generell mit steigendem Lebensalter zunimmt, so steht dies im Widerspruch zur Tatsache, daß die Heiratswahrscheinlichkeit über den Lebensverlauf ansteigt. Dieser Widerspruch ließe sich lösen, wenn man zugesteht, daß auch Frauen zu einem gewissen Anteil durch instrumentelle Aktivitäten zur Unterhaltssicherung und damit zur Familiengründung beitragen.

Als nächstes stellt sich die Frage, warum die Heiratschance von Männern und Frauen in der Mitte des zweiten Lebensjahrzehnts ihr Maximum erreicht. Während man bei den Männer sagen kann, daß in diesem Lebensabschnitt die Lösung der instrumentellen Aufgabe der Existenzsicherung weitgehend abgeschlossen ist, führt dieses Argument bei den Frauen nicht weiter. Für Frauen gilt das Gegenteil: Hinsichtlich ihrer Heiratswahrscheinlichkeit müßte man sagen, daß in diesem Alter am wenigsten Möglichkeiten zur Beteiligung an instrumentellen Aktivitäten vorhanden sind. Hier besteht die Gefahr einer zirkulären Schlußfolgerung, wenn man sagt, daß in dieser Lebensphase die Erfüllung expressiver Aufgaben am stärksten ist. Das sozialstrukturelle Argument liefert nur bei Männern eine schlüssige Erklärung. Darüber hinaus läßt sich damit nicht erklären, warum die Heiratschancen der beiden Geschlechter sich ab einem bestimmten Alter wieder angleichen.

1.3.4 Demographisch-strukturelle Erklärung

Als demographisch-strukturell möchte ich jenen Erklärungsansatz bezeichnen, der auf der Vorstellung eines "Heiratsmarktes" bzw. eines "Heiratspools" basiert. Unter dieser Perspektive ist der Zeitpunkt der Heirat im wesentlichen Ausdruck der "Suchzeit", die benötigt bzw. investiert wurde, um den gewünschten Partner innerhalb des zur Verfügung stehenden Partnerpools zu finden. "Lebensalter" wird als Indikator für die Dauer der Suche verwendet. Die Dauer der Suchzeit ist in erster Linie Ausdruck

des Umfangs des Heiratsmarktes, wenngleich noch weitere
moderierende Aspekte eine Rolle spielen (vgl. BECKER 1973). Da
der Heiratspool sich durch Eintritte und Austritte potentieller
Partner im Laufe der Zeit verändert, verändert sich auch die
Heiratschance. Zunächst wächst die Heiratswahrscheinlichkeit,
weil immer mehr Männer und Frauen als potentielle Heiratspartner
den Heiratspool erweitern. Ab einem bestimmten Zeitpunkt wird die
Heiratswahrscheinlichkeit kleiner. Dieser Zeitpunkt ist dadurch
gekennzeichnet, daß die Zahl der Austritte durch eine erfolgte
Heirat die Zahl der Eintritte in den Heiratsmarkt überwiegt. Die
Gelegenheiten, einen passenden Partner zu finden, werden im Laufe
der Zeit immer geringer, weil die Zahl der potentiellen Partner
ständig schrumpft.

Das Konzept eines dynamischen Heiratspools erscheint also
geeignet, die glockenförmige Verteilung der Heiratschance vom
Anstieg bis Abfallen der Werte zu erklären. Warum aber brauchen
Frauen weniger Zeit, bis sie einen Partner finden? Warum sind
ihre Heiratschancen in jungen Jahren, sprich: bei kürzeren
Suchzeiten größer und bei langen Suchzeiten genauso groß wie bei
Männern? Auch wenn man die Tatsache berücksichtigt, daß der
Heiratsprozeß bei Männern später einsetzt, bleibt die
Geschlechtsspezifizität des Heiratsprozesses erhalten.

Im Schaubild 2a haben wir den Prozeßanfang auf den Zeitpunkt der
gesetzlichen Heiratsmündigkeit gesetzt, denn vor diesem Zeitpunkt
ist eine Heirat gesetzlich nicht zugelassen. Aber auch in diesem
Fall ist für Frauen die Wahrscheinlichkeit höher, früher einen
Partner zur Familiengründung zu finden. Um hier eine schlüssige
Antwort im Rahmen des Konzeptes vom Heiratspool geben zu können,
müßte man die Geschlechterproportion in einem geographisch
lokalisierbaren Heiratspool in Betracht ziehen. Eine größere
Heiratschance der Frauen ließe sich dann damit erklären, daß der
Anteil der Frauen im Heiratspool größer ist als der von Männern.
Zwei Gründe lassen sich für das Zustandekommen dieser
quantitativen Relation anführen: a) Der Männerüberschuß unter den
Neugeborenen wird nicht durch eine erhöhte Sterblichkeit der
Knaben ausgeglichen. b) Die Frauen scheiden aus dem Heiratsmarkt
der Gleichaltrigen früher aus, weil sie Männer anderer Jahrgänge

heiraten. Dies bedeutet aber, daß wir die Heiratsprozesse von
Männern und Frauen ein und derselben Kohorte nicht als Ausdruck
von Suchzeiten im Heiratspool interpretieren können. Wenn man die
Altersverteilung der Heiratswahrscheinlichkeit unter Rückgriff
auf das Konzept des dynamischen Heiratspools umfassend erklären
will, muß man seinen lokalen Bezug, seinen quantitativen Umfang
und seine Geschlechterproportion kennen (vgl.
MODELL/FURSTENBERG/STRONG 1978).

1.3.5 Schlußfolgerung

Wir haben oben mehrere Möglichkeiten diskutiert, die zur
Erklärung des Befundes einer nicht-monotonen Altersverteilung der
Heiratswahrscheinlichkeit herangezogen werden können. Dies
erfolgte nicht unter der Zielsetzung, ein alle Aspekte
integrierendes Erklärungsmodell aufzustellen; es sollten nur die
spezifischen Erlärungsmechanismen der einzelnen Ansätze
analytisch herausgestellt werden. Deshalb wurden theoretische
Bezüge, die an verschiedenen Stellen der Ansätze sichtbar wurden,
nicht weiter verfolgt. Es läßt sich aber festhalten, daß eine
genuine, epigenetische Erklärung der Altersvariation der
Heiratswahrscheinlichkeit und ihrer Geschlechtsdifferenzierung
mit einer sozialstrukturelle Erklärung kontrastiert, die diese
Altersvariation als Reflektion sozialer Ungleichheit betrachtet.
In den Fällen, in denen wir eine genuin biologisch und/oder
sozial bedingte Altersabhängigkeit zugrunde legten, wurde
explizit auf die altersbegrenzte Differenzierung der
Heiratsprozesse von Männern und Frauen hingewiesen. Dies heißt
aber, daß wir prinzipiell in Rechnung stellen müssen, daß
epigenetische und sozialstrukturelle Bedingungen im Prozeß der
Ausformung des Heiratsprozesses interagieren. Im folgenden soll
der Einfluß sozio-ökonomischer Bedingungen, insbesondere des
Bildungs- und des Erwerbssystems auf die Heiratschance und
-bereitschaft beschrieben werden.

1.4   Familiengründung und Sozialstruktur

1.4.1 Allgemeine Konzeptualisierung

Einer der einflußreichsten Beiträge über die Art und Weise, wie der Zeitpunkt der Familiengründung von der Sozialstruktur einer Gesellschaft bestimmt wird, stammt von GOODE (1963) und beherrschte lange Zeit die Sichtweise in der deutschen Familiensoziologie:

"In a conjugal system, the youngsters must now be old enough to take care of themselves i.e. they must be as old as the economic system forces them to be in order to be independent at marriage. (alternative solutions also arise: some middle-class youngsters may marry upon the promise of support from their parents while they complete their education). Thus if the economic system changes its base, e.g. from agriculture to industry the age at marriage may change" (GOODE, 1963:8).

Diese ebenso knappe wie unpräzise Bestimmung der sozialen Bedingungen, die den Zeitpunkt der Eheschließung und Familiengründung beeinflussen, übernahm Rene KÖNIG (1974) in folgender Weise:

"..daß durchschnittlich das Eheschließungsalter - auf weite Sicht gesehen - von der vorwaltenden Wirtschaftsform abhängig ist, d.h. die Heiratenden sind in den verschiedenen Sozialsystemen so alt, wie die Wirtschaft es verlangt, um einen Menschen wirtschaftlich unabhängig zu machen" (KÖNIG, 1974:272).

Im Zentrum dieser These steht der Begriff der ökonomischen Unabhängigkeit, wenn auch bei GOODE der Begriff der Selbständigkeit allgemeiner gefaßt ist: "In a conjugal system, the youngsters must be ... old enough to take care of themselves...".

NEIDHARDT nimmt diesen allgemeinen Aspekt im Begriff der
"sozialen Reife" auf, den er funktional definiert: eine Person
ist sozial reif, wenn sie die von ihr erwarteten sozialen
Aufgaben bewältigen kann. Erst nach Erreichen der sozialen Reife
besteht für sie die Möglichkeit, eine Familie zu gründen.
Normativ gesehen sind die Aufgaben der Männer im Außenbereich der
Kernfamilie angesiedelt. Um die "Ernährerrolle" angemessen
erfüllen zu können, benötigen sie aber längere Ausbildungs- und
Sozialisationszeiten, die einen Aufschub ihrer sozialen Reife und
damit ihrer Familiengründung bedeuten. Frauen erreichen ihre
soziale Reife schneller, weil sie die ihnen zugewiesenen
familialen Funktionen schon in den ersten Lebensjahren in der
Herkunftsfamilie einüben (NEIDHARDT, 1974:36).

Ob eine Familiengründung spät oder früh erfolgt, hängt also
davon ab, welcher Aufwand damit verbunden ist, die soziale Reife
zu erlangen. Für die Männer bedeutet dies, daß ihre
Familiengründung hauptsächlich davon abhängt, über welchen
Bildungsweg sie ihr Erwerbsleben beginnen. Je nach Intensität der
sozialisatorischen und qualifikatorischen Vorbereitung zur
Berufsfähigkeit schieben Männer den Zeitpunkt ihrer
Familiengründung länger oder kürzer auf. Da bei den Frauen die
Entwicklung der sozialen Reife auf Erfahrungen in der
Herkunftsfamilie basiert, kann die Variation ihres Zeitpunktes
der ersten Familiengründung als Ausdruck von unterschiedlichen
sozialisatorischen und strukturellen Merkmalen der
Herkunftsfamilie interpretiert werden. Der allgemeine und
berufliche Bildungsweg im Übergang ins Berufsleben spielt für den
Zeitpunkt der Familiengründung eine geringere Rolle als bei
Männern. Da sich die gesellschaftliche Konstruktion des Zugangs
zur Familie für Männer und Frauen unterschiedlich gestaltet, ist
zu erwarten, daß sich dieses Phänomen zum einen in der
Heiratswahrscheinlichkeit der Frauen, zum anderen auch in ihrer
geringeren Beteiligung im Bildungssystem bemerkbar macht.

Dies gilt in zweifacher Weise: a) Frauen sind in geringerem Maße
am Berufszugangssystem beteiligt. b) Wenn Frauen eine
Qualifikation erreichen könnten, die mit der Qualifikation von
Männern vergleichbar ist, hätte dies in ihrem Fall einen
stärkeren negativen Effekt auf die Familiengründung.
Zusammenfassend kann man sagen, daß für NEIDHARDT die Möglichkeit
zur Berufstätigkeit und damit die Möglichkeit zur ökonomischen
Unabhängigkeit der aussungßchlaggebende Faktor der sozialen Reife
und damit der Familiengründung ist. Dies gilt jedoch nur für die
Männer. Rene KÖNIG hingegen hebt einen anderen Aspekt hervor: er
betrachtet die ökonomische Basis der Unabhängigkeit als
Determinante der Familiengründung. Der Einfluß dieser
Strukturvariablen betrifft sowohl Frauen als auch Männer und ist
im wesentlichen Ausdruck des wirtschaftlichen Wohlstandes einer
Gesellschaft. Damit wird jedoch nicht außer acht gelassen, daß
eine lange Qualifizierungsphase eine heiratsaufschiebende Wirkung
hat. KÖNIG (1974) geht davon aus, daß eine Frau mit höherer
Qualifikation nach dem Abitur eine Universität besucht, ein
Studium absolviert und erst danach, wenn überhaupt, zu einem
späten Zeitpunkt heiratet.

Die Beweggründe sind dabei unklar. Wird auf eine
Familiengründung verzichtet, weil eine höhere Berufsposition
angestrebt wird oder die durch die ausgedehnte Berufszugangsphase
erzwungene Kinderlosigkeit in der Karriere kompensiert wird.
Darüber hinaus bleiben Frauen aus höheren sozialen Schichten mit
hohem Ausbildungsstand deshalb eher unverheiratet, weil es ihnen
kaum möglich ist "hinaufheiraten". Da ihr eigener Status durch
den Status des Ehemannes definiert wird, müssen sie befürchten,
daß ihr eigener hoher Status durch eine Heirat sinkt.

MEYER/SCHULZE (1983) greifen erneut das Argument auf, daß
Frauen, die durch ein hohes Ausbildungsniveau mit beruflichen
Karrieremöglichkeiten in ökonomischer Hinsicht selbständig
geworden sind, in geringerem Maße bereit sind, eine Familie zu
gründen (MEYER/SCHULZE 1983).

Sie gehen davon aus, daß in der Familie die generative,
konsumptive und emotionale Regeneration des Menschen möglich ist

und überwiegend auch von Frauen erwartet wird. Je größer der Raum
ist, den die Erwerbstätigkeit in der Lebensperspektive von Frauen
einnimmt, desto größer ist ihr Regenerationsbedürfnis und desto
weniger wollen und können sie die psychische Regeneration von
Mann und Kindern übernehmen. Dabei spielt es weiterhin eine
Rolle, daß sie nicht mehr gezwungen sind, ihre Versorgung durch
eine Eheschließung zu sichern. Diese Umstände führen dazu, daß
Frauen mit hoher Qualifikation Ansprüche auf egalitäre
Beziehungsformen in einer Partnerschaft stellen. Die
nicht-eheliche Lebensgemeinschaft scheint ihnen zur
Verwirklichung ihrer Vorstellungen vielversprechender und daher
verzichten sie auf die Eheschließung.

1.4.2 Zusammenfassung: Berufszugang und Familiengründung

Aus der obigen Diskussion läßt sich folgende zentrale These
bezüglich des Einflusses des Berufszugangs auf den Zeitpunkt der
Familiengründung ableiten: Je anspruchsvoller der Berufszugang
ist, um so stärker sinkt die Wahrscheinlichkeit der
Familiengründung.

Empirische Hinweise dafür findet man in Auswertungen des
Statistischen Bundesamtes (1964), bei JÜRGENS (1973:115) und bei
RÜCKERT/LENGSFELD/HENKE (1979:83). Da der Berufseintritt für
Einkommen und Entwicklung des Berufsweges die Weichen stellt
(vgl. BLOSSFELD 1985) gelangen Frauen durch die Wahl der
beruflichen Ausbidung und den Einstiegsberuf weitaus mehr in
"Sackgassen"berufe ohne Aufstiegschancen (ROSENFELD 1980, KANTER
1977, MARINI 1980), mit schlechterer Arbeit und geringerem
Einkommen (LAPPE 1985). Angesichts der Tatsache, daß Frauen durch
ihre schlechtere Ausstattung mit Humankapital geringere Chancen
auf dem Arbeitsmarkt haben, sind sie hinsichtlich ihrer
Versorgungs- und Selbstverwirklichungsinteressen auf die Familie
zurückgelenkt. Allerdings gibt es hierzu widersprüchliche
Ergebnisse. FEATHERMAN/HAUSER (1976), MCCLENDON (1976),
TREIMAN/TERREL (1975) kommen zu dem Schluß, daß der Statuserfolg
im Laufe der Berufskarriere für Männer und Frauen im wesentlichen
der gleiche ist. SEWELL/HAUSER/WOLF (1980) hingegen fanden bei

Abgängern eines Jahrganges der High School, daß Frauen ihre
Berufswege mit einem höheren Berufsstatus beginnen als Männer,
von denen ein geringerer Teil den Berufsweg mit einem Büroberuf
beginnt. Viele Männer treten zwar als un- oder angelernte
Arbeiter ins Erwerbsleben, aber der weitere Verlauf ihres
Berufslebens ist durch einen Aufstieg gekennzeichnet. Im
Berufsweg von Frauen ist dagegen eher ein Abstieg oder eine
Stagnation des beruflichen Status zu beobachten.

Da Frauen keine Chancen zur Mobilität haben, müssen sie ihre
familialen Pläne früher als Männer realisieren. Oder verhält es
sich umgekehrt? Verzichten sie auf ihre Aufstiegschancen, damit
sie ihre familialen Pläne früher verwirklichen können?

Ein anpruchsvoller Berufszugang wirkt sich bei Frauen weniger
auf den Zeitpunkt der Familiengründung aus als bei Männern. Schon
in den frühen Lebensjahren entwickeln Kinder ein ziemlich klares
Verständnis von geschlechtsspezifischen Rollen und artikulieren
ihre Berufswünsche in geschlechtsspezifischer Form. Schon vor dem
Übergang ins Erwerbssystem ist für die Jungen klar, daß sie den
Erwachsenenstatus hauptsächlich über berufliche Leistungen
erreichen können. Sie benutzen alle verfügbaren Ressourcen, um
berufliche Erfolge zu erreichen. Da bei Mädchen berufliche
Erfolge weniger belohnt werden, mobilisieren sie die gleichen
Ressourcen in einem geringeren Ausmaß als Männer. (MARINI
1978:747). Beim Übergang in den Arbeitsmarkt werden Konflikte
zwischen familialen und beruflichen Rollen vom Heranwachsenden
geschlechtsspezifisch gelöst. Wenn Mädchen spät heiraten und
wenige Kinder haben wollen, entwickeln sie berufliche Pläne mit
hohem Status, und wählen eher geschlechtsuntypische Berufe.
Familiale Rollenvorstellungen sind für die Berufswünsche der
Jungen hingegen irrelevant. Es wird deutlich, daß für Männer
Arbeit und Familie getrennte Handlungssphären repräsentieren
(ANESHENSEL/ROSEN 1980).

Eine Folge der Rollensegregation ist, daß die oben genannten
sozialstrukturellen Bedingungen der Familiengründung für Männer
und Frauen unterschiedliche Bedeutung erhalten. Der Einfluß der
Ausbildungsdauer auf das Heiratsalter ist bei Frauen größer als

bei Männern. Dies unterstützt die These, daß Familie und individuelle berufliche Leistung für Frauen komplementär sind, für Männer jedoch handelt es sich um getrennte Welten (MARINI 1978). Die Studie von RINDFUSSUß et al. (1980) zeigt, daß auch der Zeitpunkt der Geburt des ersten Kindes durch die Ausbildungsinvestitionen vorausgesagt werden kann. Allerdings kamen WAITE/SPITZE (1981) und GOLDSCHEIDER/WAITE (1985) zu einem anderen Ergebnis: die Heiratswahrscheinlichkeit wurde um so größer, je länger der Schulbesuch dauerte. Im Gegensatz dazu ergab sich in der Studie von RINDFUSS (1980) und MARINI (1978) ein negativer Effekt der Ausbildungsdauer auf das Heiratsalter. Es ist aber zu vermuten, daß der positive Zuammenhang von Schuldauer und Heiratswahrscheinlichkeit, wie er in den einzelnen Altersgruppen gefunden wurde, nur den Unterschied von Personen mit beendeter Ausbildung gegenüber Personen, die sich in der Ausbildung befinden, widerspiegelt.

WAITE/SPITZE (1981) erklären den positiven Einfluß der längeren Ausbildungsdauer folgendermaßen: Die höhere Ausbildung führt dazu, daß die Attraktivität und Verfügbarkeit der Frauen als Ehefrauen zunimmt (691). Die Tatsache, daß erwerbstätige Frauen eine höhere Heiratswahrscheinlichkeit als nicht-erwerbstätige Frauen haben, steht ebenfalls in Widerspruch zur obigen Ableitung, daß diese Frauen aufgrund ihrer ökonomischen Unabhängigkeit nicht zu heiraten brauchen. Die post-hoc Erklärung von WAITE/SPITZE (1981) verweist auch hier darauf, daß die Attraktivität der Frauen durch ihre Berufstätigkeit steigt, zumal sie mehr Kontakt zu potentiellen Partnern haben als Frauen, die nicht in den Arbeitsmarkt integriert sind.

1.5 Kohortendifferenzierung und Familiengründung

Die oben aufgeführten Annahmen über die Bedingungen der Familiengründung sind i.d.R. ahistorisch formuliert. D.h. sie zielen darauf ab, die Familiengründung als einen sozialen Tatbestand zu beschreiben, der ein dauerhaftes Merkmal westlicher Industriegesellschaften darstellt. Es wäre aber verfehlt, die westliche Industriegesellschaft als eine Gesellschaftsform zu betrachten, die sich im Gleichgewicht befindet. Eines ihrer

wesentlichen Merkmale betrifft gerade den rapiden Wandel von
Normen und Sozialstrukturen im Laufe ihrer historischen
Entwicklung. Das Stichwort "Industrialisierung" mag hier als
Hinweis auf die langfristige Entwicklungsbewegung der westlichen
Gesellschaft genügen. Technische Innovationen und
Produktivitätsfortschritte verändern die wirtschaftliche Basis
des familialen Systems in einer Wirtschaftsorganisation, deren
konstituierendes Element der individuelle Akteur ist, der seinen
Lebensunterhalt durch das eigene Arbeitsvermögen und durch das
darin investierte Bildungs- und Qualifikationskapital bestreiten
kann (vgl. ROSENBAUM, 1978). In der Bundesrepublik Deutschland
expandierte nach dem Zweiten Weltkrieg und der Wiederaufbauphase
auch durch verstärkte Aktivitäten des Staates der tertiäre Sektor
und die berufliche Bildung gewann größere Bedeutung im
Lebensverlauf der Individuen. Die zentrale Bedeutung von
Qualifikation und Wissen als direktes und indirektes
Produktionsmittel ist ein markantes Kennzeichen der
fortgeschrittenen Industriegesellschaft.

Wie veränderte sich die Familienstruktur und die Bedingungen der
Familiengründung im Gefolge dieser sozio-ökonomischen
Entwicklung?

Die zeitgenössische Familiensoziologie liefert widersprüchliche
und diffuse Beiträge zu dieser Frage. Die Gründe dafür sind
einerseits darin zu sehen, daß nicht auf eine genau spezifizierte
historische Periode Bezug genommen, sondern sehr global auf
Entwicklungsstufen der westlichen Gesellschaft abgehoben wird.
Andererseits gehen sie von unterschiedlichen Annahmen über die
soziale Basis der Familiengründung aus. "Unterentwickelte
Gesellschaften, die wenige Fertigkeiten benötigen, weisen ein
niedrigeres Heiratsalter als industrielle Gesellschaften auf. In
der fortgeschrittenen Industriegesellschaft ändert sich das
wieder" (KÖNIG, 1974:245).

Dieser Aussungßage läßt sich entnehmen, daß das Heiratsalter in
der fortgeschrittenen Industriegesellschaft wieder sinkt. Es wird
aber auch betont, daß in dieser Gesellschaftsform die
Familiengründung tendenziell aufgeschoben wird, denn aufgrund

ihres höheren Ausbildungsniveaus und ihrer verstärkten sozialen
Aufstiegstendenz entwickeln viel mehr Frauen höhere berufliche
Erwartungen (KÖNIG, 1974:259). In diesem Prozeß werden sie
zunehmend unabhängiger vom Mann als Versorger, weil ihnen selbst
in höherem Maße alternative Erwerbsmöglichkeiten zur Verfügung
stehen (KÖNIG,1974,377).

NAVE-HERZ (1967) geht ebenfalls von der Annahme aus, daß
Veränderungen im Bildungssystem zum Aufschub der Familiengründug
geführt haben. Als Bestätigung dieser Annahme wird angeführt, daß
das Heiratsalter "früher" (im Bern des 17. Jahrhunderts)
durchschnittlich niedriger war, weil sowohl Männer und als auch
Frauen einen niedrigeren Ausbildungsstand hatten (s.272).

MEYER/SCHULZE (1983) nehmen hingegen auf neuere Entwicklungen
Bezug, um die Tatsache zu erklären, daß Frauen in den siebziger
Jahren eine vergleichsweise geringere Bereitschaft zur Ehe
zeigten und stattdessen zu nicht-ehelichen Lebensgemeinschaften
tendierten. Als Ursache dafür betrachten auch sie die
verbesserten Ausbildungsmöglichkeiten und Berufschancen der
Frauen in der fortgeschrittenen Industriegesellschaft. In der
modernen Gesellschaft ermöglicht die ökonomische Selbständigkeit
den Frauen, sich von der einseitigen familialen Abhängigkeit zu
befreien. Frauen distanzieren sich zunehmend von der einseitigen
Festlegung auf Funktionen der Regeneration der Familienmitglieder
und wenden sich stärker anderen Formen des Zusammenlebens zu.

Das Hauptproblem dieser und ähnlicher Analysen ist darin zu
sehen, daß bei der Beurteilung des sozialen Wandels der
Familiengründung nicht zwischen Kohorten- , Perioden- und
Alterseffekten differenziert wird. Werden diese Aspekte des
sozialen Wandels nicht deutlich voneinander getrennt, so nehmen
die Aussagen über die historische Veränderung der Familienbildung
sehr leicht die Form von "früher-jetzt"-Aussagen mit beliebigem
Gehalt an. Wenn man aber den historischen Kontext der
Wirtschafts- und Sozialstruktur der westlichen Gesellschaft in
seinem Einfluß auf die Familiengründung nachzeichnen will, so muß

man Ausmaß und Richtung der Differenzierung in sukzessiven
Kohorten betrachten. Der historische Kontext kann sich in
zweifacher Hinsicht auf die Entscheidung zur Familiengründung
auswirken: 1) die wirtschaftlichen und/oder sozialen Bedingungen
eines bestimmten historischen Zeitpunktes oder Intervalles
beeinflussen gleichermaßen alle Personen, und zwar unabhängig von
ihrem Lebensalter (Periodeneffekt). 2) der Effekt spezifischer
historischer Umstände kommt nur bei Personen in einem bestimmten
Lebensalter zum Tragen. In diesem Fall handelt es sich um einen
Kohorteneffekt, da bei gegebenem kritischen Alter der historische
Einfluß sich nur bei spezifischen Jahrgängen zeigen wird.

Weiterhin kann man die Auswirkungen der historischen Umstände
danach unterscheiden, ob sie einen prägenden Einfluß oder nur
einen vorübergehenden Einfluß bedeuten. D.h. man kann fragen, ob
ein Aufschub zur Fixierung wird, also zum Verzicht wird, oder ob
eine Kompensation, d.h. ein Nachholen z.B. der Familiengründung
zu einem späteren Zeitpunkt möglich ist (CARLSON/KARLSON 1970).
Folgende Aspekte sind zu berücksichtigen, wenn man eine
sozialhistorische Entwicklung und ihre Auswirkungen auf die
Bereitschaft oder Möglichkeit zur Familiengründung in der Sequenz
mehrerer Kohorten beschreiben will: 1. Die Angabe aller
potentiellen Einflußfaktoren. Der historische Einfluß entspricht
dem Einfluß dieser Mechanismen in ihrer historisch spezifischen
Konstellation. 2. Die Auswahl der aufeinander folgenden Kohorten.
3. Die Angabe der kritischen Lebensphase, um die spezifische
historische Konstellation der Einflußfaktoren zu benennen.

Die Auswahl der Kohorten ist in der vorliegenden Arbeit durch
das Design der Studie bedingt. Sie war im Hinblick auf
Veränderungen im Bildungs- und Berufssystem in der westdeutschen
Gesellschaft nach dem Zweiten Weltkrieg getroffen worden. Im
einzelnen handelt es sich um die Jahrgänge 1929-31, 1939-41 und
1949-51. Mangels hinreichender empirischer Fundierung ist die
Definition der kritischen Lebensphase schwierig. Je nachdem, ob
man sich an entwicklungspsychologischen Adoleszenztheorien oder
an psychoanalytischen bzw. Lerntheorien orientiert, ergeben sich
unterschiedliche Intervalle als kritische Lebensjahre. Die
Bestimmung dieser Lebensjahre hängt weiterhin davon ab, welches

Zielverhalten angesprochen wird. Die Erläuterungen über die Bedingungen der Altersabhängigkeit haben verdeutlicht, daß sich eine Lebensphase eingrenzen läßt, in der langfristige Entscheidungen in bezug auf Partnerwahl und Familiengründung von Bedeutung sind. Diese Sichtweise ermöglicht es, die prägende Wirkung von Erfahrungen in Kindheit und Adoleszenz zunächst außer acht zu lassen.

Bei der Festlegung der kritischen Heiratsphase orientiere ich mich einerseits am gesetzlichen Mindestalter bei erster Heirat (für Frauen das 16. und für Männer das 18. Lebensjahr). Hinsichtlich der oberen Grenze der heiratskritischen Lebensphase gibt es höchstens Anhaltspunkte aus Untersuchungen über Altersnormen, die sehr stark je nach der historischen Bezugsperiode variieren (MODELL 1978). Wenn man aber als weiteren Aspekt einbezieht, in welchen Altersgruppen der soziale Wandel der Heiratswahrscheinlichkeit von Frauen und Männern besonders deutlich erfolgte (vgl. PAPASTEFANOU 1986), dann bietet sich folgendes Intervall als kritische Lebensphase der Familiengründung an: bei Frauen: vom 17. bis 24. Lebensjahr mit Schwerpunkt beim 22. Lebensjahr bei Männern: vom 19. bis 26. Lebensjahr mit Schwerpunkt beim 24. Lebensjahr.

Diese Entscheidung führt zu folgenden für die Familiengründung kritischen historischen Zeitabschnitten: Kohorte A: 1947/49 bis 1954/56; Stichpunkt: 1952. Kohorte B: 1957/59 bis 1964/66; Stichpunkt: 1962. Kohorte C: 1967/69 bis 1974/76; Stichpunkt: 1972. Wir müssen nun nach den sozialen und wirtschaftlichen Indikatoren fragen, die aufgrund der oben entwickelten Einflußmechanismen in diesen Zeitabschnitten von Bedeutung sind.

Als Indikatoren könnten folgende gesamtgesellschaftliche Merkmale dienen:

- wirtschaftlicher Wohlstand

- berufliche Chancen / Arbeitsmarkt

- allgemeine Bildungschancen

- berufliche Bildungschancen

- sexuelle Liberalität

- Verhütungsmöglichkeiten

Auf eine genaue quantitative Beschreibung der genannten
Zeitabschnitte anhand dieser Indikatoren soll hier verzichtet
werden. Es soll innerhalb für die Absicht der vorliegenden Arbeit
genügen, a priori eine skizzenhafte Deskription zu geben und die
weitere detaillierte Beschreibung dieser historischen Phasen in
die nachträgliche Interpretation der bereinigten
Kohortenunterschiede aufzunehmen.

Die drei Zeitabschnitte kennzeichnen drei wichtige Phasen der
Entwicklung der westdeutschen Gesellschaft: a) Wiederaufbau, b)
Restauration und Wohlstand, a) Reformen im Wohlstand.

Wiederaufbau

Die Zeit unmittelbar nach dem Krieg ist durch einen schlechten
Arbeitsmarkt, zerstörter Infrastruktur des Bildungswesens und der
traditionellen Arbeitsteilung gekennzeichnet. Formale
Ausbildungsmöglichkeiten wie Lehre, Fachschulen und
Universitätsausbildungen waren beschränkt. Die Ertragskraft der
Wirtschaft war niedrig und durch geringe Löhne gekennzeichnet.
Durch den kriegsbedingten "Ausfall" der Männer waren Frauen
verstärkt auf eigentlich männlich bestimmte Arbeitsplätze und auf
die selbständige Bestreitung ihres Lebensunterhaltes angewiesen.
Heiratsfähige Männer waren gefallen oder in Kriegsgefangenschaft.

Restauration und Wohlstand: "die Goldenen Fünfziger und Sechziger
Jahre"

Der erfolgte Wiederaufbau und die Modernisierung der
Wirtschaftsstruktur hatte zur Folge, daß die berufliche
Qualifikation einen größeren Stellenwert erlangte. In einer
expandierenden Wirtschaft gewannen die berufliche Bildung und die

Berufskarriere einen eigenen Stellenwert. Mit wachsendem
Wohlstand ging der Glaube an die berechenbare, materiell
gesicherte Zukunft einher. Diese wirtschaftliche Entwicklung
rekonstituierte die traditionelle Arbeitsteilung. Soziale
Kontrolle im Sinne sexueller Normierung und Traditionalisierung
gewannen wieder an Bedeutung.

Reformen im Wohlstand: Bildungsexpansion

Folgende historische Ereignisse fallen in diese Zeit: die
Rezession 1967/68, die Große Koalition, der Übergang zur
sozial-liberalen Koalition und die Studentenbewegung mit dem
Anfang der sexuellen Liberalisierung. Besondere Aufmerksamkeit
erhielten die Forderung nach größeren Bildungschancen für Kinder
vom Land, für Arbeiterkinder und für Mädchen. Durch Absolvierung
einer beruflichen Ausbildung verbessern sich insbesondere die
beruflichen Chancen der Frauen, zumal sie ihren Berufseintritt in
einer Zeit der Expansion von staatlichen Ausgaben in Gesundheit
und Bildung, und der Dienstleistungsbranche insgesamt vornahmen.
Frauen finden ihre Berufe vermehrt in qualifizierten
Dienstleistungs- und Verwaltungsberufen. Allerdings bedeutet die
Ausweitung dieser Wirtschaftsbereiche auch, daß die
Erwerbstätigkeit von Frauen sich verstärkt in typischen
Frauenberufen konzentrieren wird (BLOSSFELD 1985b).

Die sexuelle Liberalisierung folgt verzögert der Diffusion
oraler Verhütungsmittel. Allerdings kann man davon ausgehen, daß
noch in der Kohorte 1949-51 nur in geringem Maße von der Pille
Gebrauch gemacht wurde. Es öffnet sich eine "Schere" zwischen
wachsender sexueller Liberalisierung und noch geringer
Verfügbarkeit von Verhütungsmitteln. Aufgrund des Babybooms nach
den Zweiten Weltkrieg stehen diesen Frauen mehr Männer auf dem
Heiratsmarkt gegenüber als das für Frauen in der Nachkriegszeit
der Fall war. Die Männer erleben einen "marriage squeeze".

Insgesamt kann man die Entwicklung der westdeutschen
Gesellschaft als eine lineare Entwicklung zur tertiarisierten
Industriegesellschaft sehen (BLOSSFELD 1985b), in der die
Geschlechtsdifferenzierung der Familiengründung durch Abbau von

Bildungsungleichheit zunehmend nivelliert wird. Diese These kann aufgrund der Komplexität des sozialen Wandels keiner direkten Überprüfung unterzogen werden. Außerdem stellt die Linearisierung von Entwicklungstrends eine unzulässige Hypostasierung dar. Der historische Kontext muß für die aufeinanderfolgenden Kohorten einzeln spezifiziert werden. Da aber die jeweiligen historischen Situationen zu komplex sind und wir in der vorliegenden Analyse quantitative Indikatoren dieser Kontexte nicht berücksichtigen, soll auf die Explikation der kohortenspezifischen Einflüsse des sozialhistorischen Kontextes auf die Familiengründung an dieser Stelle verzichtet werden. In der vorliegenden Arbeit dienen die Kohorten gleichsam als Kontrollgruppen bei der Überprüfung der Hypothesen über den Zusammenhang von Bildungssystem und Familiengründungsprozeß. Die posthoc-Betrachtung der historischen Differenzierung des Familiengründungsprozesses bietet dann eine Möglichkeit zur Generierung präziserer Hypothesen.

2. Empirische Analyse

2.1 Daten und Methode

2.1.1 Beschreibung der Daten

Die folgende empirische Analyse beruht auf den Daten des Teilprojektes "Lebensverlauf und Wohlfahrtsentwicklung", das im DFG-Sonderforschungsbereich 3 "Mikroanalytische Grundlagen der Gesellschaftspolitik" durchgeführt wurde. In dieser Studie wurden 2 171 Männer und Frauen befragt. Sie repräsentieren die deutsche Wohnbevölkerung der Jahrgänge 1929-31, 1939-41 und 1949-51, die zum Befragungszeitpunkt in Westdeutschland lebten. Eine detaillierte Beschreibung der Stichprobe gibt WIEDENBECK (1982). Die Erhebung erstreckte sich von Oktober 1981 bis Mai 1983 (vgl. BRÜCKNER u.a. 1984). Mit Hilfe eines standardisierten Interviews wurden die Lebensverläufe dieser Personen in bezug auf die Lebensbereiche Herkunftsfamilie, Ausbildung, Beruf, Familie, Wohnung, Gesundheit und politische Partizipation erfaßt. Bei diesen Angaben handelt es sich um Rückerinnerungsdaten. Die befragten Personen wurden gebeten, die Ereignisse in den verschiedenen Bereichen mit möglichst genauen Zeitangaben

anzugeben. Das Ausmaß der Zuverlässigkeit der retrospektiven
Daten wurde nur z.T. untersucht. Die bislang durchgeführten
Reliabilitätsstudien sprechen jedoch im allgemeinen dafür, daß
die Daten den üblichen Anforderungen an Zuverlässigkeit
entsprechen, das üblicherweise für Querschnittsuntersuchungen
berichtet wird (BLOSSFELD 1985a). Ein Vergleich mit den amtlichen
Zahlen ergab, daß die Altersverteilung der Heiratsraten von
Frauen der in der vorliegenden Studie ausgewählten Kohorten eine
verläßliche Rekonstruktion des Heiratsprozesses in der Population
darstellt (PAPASTEFANOU 1985).

An dieser Stelle soll gezeigt werden, daß durch die Auswahl
dieser spezifischen Kohorten, bei den Veränderungen im
Ausbildungs- und Berufssystem im Vordergrund standen, eine ganz
besondere Phase der historischen Entwicklung des Familiensystems
herausgegriffen worden ist. Bei den Frauen trifft man auf eine
historische Periode der relativen Stabilität des
Heiratsprozesses, in der sich ihre Heiratschancen insgesamt
verbesserten (siehe Schaubild 3 und 4), bevor nach dem Jahrgang
1951 eine deutliche Umstrukturierung des Heiratsprozesses
einsetzt. Bei den Männern hingegen greifen wir mit den drei
Kohorten drei Schnitte aus einer längerfristigen
Altersnivellierung der Heiratschance heraus. Bei den Männern der
Kohorte 1949-51 können wir einen Heiratsprozeß beobachten, der
bei den Frauen erst in den Jahrgängen 1956 und später auftritt
(vgl. Schaubilder 5 und 6). Aufgrund der Tatsache, daß Männer
insgesamt später heiraten, gelangen sie - zusätzlich verstärkt
durch ihre längere sensitive Heiratsphase - sozusagen "tiefer" in
den Einfluß des historischen Kontextes der Bildungsexpansion.

2.1.2 Operationalisierung: parametrisches Modell der
Altersabhängigkeit und Codierung der Prädiktoren

Zur Überprüfung der oben aufgestellten Hypothesen wird ein
multivariates stochastisches Modell verwendet. Die Wahl dieses
Verfahrens geschieht aus drei Gründen: 1. Die abhängige Variable,
wie sie sich im Rahmen einer Lebenslaufperspektive als
Erklärungsobjekt ergibt, ist die Wahrscheinlichkeit der Heirat in
jedem Lebensmoment eines Individuums. Die angemessene

Operationalisierung dafür ist die "hazard rate". Diese wird auf
der Basis der individuellen Wartezeiten bis zum Ereignis
geschätzt und mittels der ausgewählten Kovariaten prädiziert. Als
initiierendes Ereignis benutzen wir den Zeitpunkt der Erlangung
der gesetzlichen Heiratsmündigkeit. Daher wird zur Bildung der
individuellen Ereignisrisiken die jeweilige Wartezeit von
Heiratsmündigkeit bis Eheschließung verwendet.

2. Jede Analyse sozialer Prozesse ist mit dem Problem
konfrontiert, daß ein Teil der Personen das in Frage kommende
Ereignis noch nicht realisiert hat (zensierte Fälle). Wenn man
diese Personen, für die es ja keine echte Wartezeit gibt, aus der
Analyse herausnimmt, kann dies zu einer Verzerrung der Ergebnisse
führen. Denn Bevölkerungsgruppen, die sich durch ein spätes
Heiratsalter auszeichnen, sind dann in der Stichprobe
unterrepräsentiert. Die Maximum-Likelihood-Schätzung kann dieses
Problem lösen, indem für jede dieser Personen auf der Basis der
log-logistischen Altersverteilung eine hazard rate geschätzt
wird.

3. Aufgrund der theoretischen Überlegungen müssen wir davon
ausgehen, daß ein genuiner Alterseffekt im Sinne einer
nicht-monotonen Altersverteilung der Hazardrate vorliegt. Dies
ist gewöhnlich das ausschlaggebende Argument bei der Wahl eines
parametrischen Modells der Hazardrate (SOERENSEN/SOERENSEN 1984).

Eine mit dieser Entscheidung verbundene Konsequenz wurde meines
Wissens bisher nicht diskutiert. Es handelt sich dabei um
folgenden Sachverhalt: Wenn man in einem Modell durch die Wahl
der Altersfunktion eine für die Heiratsentscheidung kritische
Lebensphase annimmt, dann muß auch eine Altersdifferenzierung der
Kovariateneinflüsse berücksichtigt werden. Es ist möglich, daß
die durch den Prädiktor indizierten Gruppen sich hauptsächlich in
dieser kritischen Heiratsphase bezüglich ihrer
Heiratswahrscheinlichkeit unterscheiden. Die wenigen Arbeiten, in
denen die altersspezifischen Faktoreinflüsse untersucht werden
(RINDFUSS et al.1985, GOLDSCHEIDER/WAITE 1985, WAITE/SPITZE 1982)
zeigen, daß sich der Einfluß von Determinanten wie

Erwerbstätigkeit und Bildungsdauer auf die
Heiratswahrscheinlichkeit in einem nicht-monotonen Verhältnis zum
Lebensalter verändert. Bei der Spezifikation der nicht-monotonen
Altersfunktion der Hazardrate orientiere ich mich an der
Untersuchung von SOERENSEN/SOERENSEN (1984), indem ich das
log-logistische Wartezeitmodell in der Parametrisierung von
PETERSON (1983) verwende (vgl. auch KALBFLEISCH/PRENTICE 1980 zur
Frage der Modellierung nicht-monotoner Prozesse).

Das Modell hat folgende funktionale Form:

$$h(t) = \frac{((1+g)\exp(b'X+g*\ln(t)))}{(1+\exp(b'X+(1+g)\ln(t)))}$$

h(t) ist die Hazardrate zum Zeitpunkt t, g ist der Effekt der
Wartezeit auf die Rate, X steht für eine Menge von exogenen
Variablen und b für ihre Effekte auf die Rate. Je nachdem, ob das
Altersgewicht g Werte größer oder kleiner 1 annimmt, wird sich
die Altersabhängigkeit des Modells auch in der Verteilungsform
ändern. Bei Werten unter 1 nimmt die log-logistiche Verteilung
eine Form an, die der Weibull-Verteilung entspricht, bei Werten
über 1 erhält sie die Form einer rechtsschiefen Glockenkurve
(vgl. hierzu auch DIEKMANN/MITTER 1984:153f).

Da die numerischen Werte der geschätzten Effekte aufgrund der
Funktion nicht unmittelbar beurteilt werden können, wurde die
mittlere Verweildauer für die relevanten Subgruppen aufgrund
ihrer geschätzten Effekte nach folgender Formel (vgl. PETERSON
1984) berechnet:

$$E(T/T<=tm) = (1/P(T<=tm))*INT(t*f(t)dt$$

E(T/T<=tm): mittlere Wartezeit bis zur Heirat, unter der
Bedingung, daß eine Person bis zum Zeitpunkt tm heiratet. In den
unten wiedergegebenen Werten von E() wurde als tm die Wartezeit
von 408 Monaten verwendet. Dies entspricht bei Frauen einem
Lebensalter von 50 und bei Männern einem Lebensalter von 52
Jahren.

P(T<=tm): Wahrscheinlichkeit, in dem Zeitraum von t0 bis tm zu
heiraten

f(t): Wahrscheinlichkeitsdichte zum Zeitpunkt t.

INT: Integral von t0 bis tm ...

Unsere theoretischen Überlegungen führten uns ferner dahin, die
Heiratswahrscheinlichkeit als von sozio-ökonomischen Merkmalen
des Berufszuganges beeinflußt zu betrachten. Da es sich dabei im
wesentlichen um kategoriale Variablen handelt, haben wir sie in
Form von dichotomen Prädiktoren in das Modell aufgenommen. Die
Definition dieser Prädiktoren ist in der Übersicht 1
wiedergegeben.

2.2. Ergebnisse

2.2.1 Die "bereinigte" Geschlechterdifferenzierung der
Heiratswahrscheinlichkeit

Die erste Modellrechnung ergibt folgende Schätzwerte für die
Unterschiede zwischen den sozialen Gruppen (siehe Tabelle 1; die
Mittelwerte der Prädiktoren sind in Tabelle A-1 im Anhang). Da
die Interpretation dieser Effekte aufgrund ihrer komplexen
Interaktion mit dem Alterseffekt nicht unmittelbar möglich ist,
haben wir sie der Verständlichkeit halber als Gruppenunterschiede
der mittleren Wartezeit bis zur ersten Heirat ausgedrückt.

Allerdings müßte man diese Betrachtung dadurch differenzieren,
daß man einen Bezug zur kritischen Lebensphase der
Familiengründung herstellt. Wenn man von einem Lebensalter
ausgeht, ab dem die Heiratswahrscheinlichkeit sehr gering bzw.
Null ist, dann bedeutet ein Aufschub der Familiengründung über
dieses Lebensalter hinaus, daß die Gelegenheit zur
Familiengründung verloren geht. Je nachdem, zu welchem Zeitpunkt
man dieses Alter ansetzt, wird eine lange Wartezeit entsprechend
größere Verzichtkonsequenzen haben. Dies wird sich besonders in
der Wartezeit niederschlagen, die man noch vor sich hat, wenn man
ein bestimmtes kritisches Lebensalter überschritten hat.

Man müßte sich fragen: Welches Lebensalter ist das kritische Alter und wie hängt es mit der Chance zusammen, noch zu heiraten? Wie viele von denen, die z.B. 25 Jahre und ledig sind, werden noch bis zum Alter von 50 heiraten? Wie lange dauert es bei ihnen durchschnittlich, bis sie sich zur Heirat entschließen?

Um die Ergebnisse überschaubar zu halten, werde ich diese Gesichtspunkte in einem zweiten Teil der Analyse aufgreifen. Im vorliegenden Teil habe ich die Zeitspanne bis zum 50. Lebensjahr als Lebensphase der möglichen Heirat ausgewählt. Innerhalb dieser Zeit heiraten die meisten Personen der vorliegenden Kohorten. Der Berufszugang hat kaum differentielle Effekte des Verzichts auf die Familiengründung (Tabelle 2). Die Wahrscheinlichkeit, bis zum 50. Lebensjahr zu heiraten, ist sehr hoch und liegt bei ca. 98 Prozent. Die einzige Gruppe, die nennenswert davon abweicht ist die Gruppe der Abiturienten mit einem Anteil von 94 Prozent. Die eigentliche Differenzierung der Heiratschance durch den Bildungsweg drückt sich darin aus, daß ein Unterschied der mittleren Wartezeit bis zur Heirat vorhanden ist (Tabelle 3). Erwartungsgemäß sind eine höhere Schulbildung und eine höhere Qualifikation damit verbunden, daß eine längere Zeit ohne eigene Familie verbracht wird.

Ein wesentlicher Aufschub der Familiengründung wird durch den Besuch eines Gymnasiums bedingt. Der Unterschied gegenüber den Hauptschülern entspricht ungefähr dem zeitlichen Unterschied von 4.5 Jahren, mit dem sie die Schule später verlassen. Es ist in diesem Zusammenhang auffällig, daß der Unterschied zwischen den Haupt- und den Realschülern sehr viel niedriger ausfällt, als man dies aufgrund des unterschiedlichen Zeitpunktes ihres Schulaustrittes erwarten würde. Die Erlangung einer akademischen Qualifikation ist damit verbunden, daß die Familiengründung im Vergleich zu einer Lehrlingsausbildung um 2.1 Jahre aufgeschoben wird. Dies entspricht wiederum der zusätzlichen Ausbildungszeit, die ein Studium im Vergleich zur Lehre mehr in Anspruch nimmt. Aber auch in den Fällen, in denen keine Ausbildung erlangt oder die Ausbildung ohne Abschluß beendet wurde, kann eine allerdings nur geringe Verzögerung der Familiengründung beobachtet werden.

Frauen heiraten etwas mehr als drei Jahre früher als Männer,
auch wenn man die Bildungs- und Qualifikationsunterschiede der
Geschlechtergruppen konstant hält. D.h. wir können den
Altersunterschied nicht als Ausdruck ihres unterschiedlichen
Berufszuganges erklären. Der Unterschied von Frauen und Männern
hinsichtlich der Wartezeit bis zur Eheschließung bleibt der
gleiche, auch wenn man die Faktoren des Berufszugangs nicht in
das Modell aufgenommen hat (Tabelle 4).

Entgegen den Erwartungen erweisen sich die Kohortenunterschiede
nicht als Effekte der veränderten Bildungsbeteiligung. Es bleiben
deutliche Kohortendifferenzen hinsichtlich des mittleren
Heiratsalters erhalten, auch wenn man die Bildungsdifferenzierung
der Population statistisch kontrolliert. Die Tendenz geht dabei
in die Richtung, daß die Wartezeit bis zur Heirat in dieser
historischen Periode kürzer wird. Diese Tendenz spiegelt eine
Entwicklung wider, die durch gleichzeitige Veränderungen im
Bildungs- und Berufssystem verdeckt wurde. Denn die
Vorverlagerung des Heiratszeitpunktes in der Abfolge dieser
Kohorten wird stärker, wenn man die Einflüsse des Berufszugangs
konstant hält.

Ein überraschendes Ergebnis besteht darin, daß durch Hinzunahme
von Kovariaten die Altersabhängigkeit stärker wird. Darauf werde
ich im Abschnitt 2.2.6 näher eingehen.

Die Wahl einer nicht-monotonen Beziehung zwischen
Heiratswahrscheinlichkeit und Wartezeit hat zur Folge, daß die
Differenzierungswirkung des Berufszugangs für jedes Lebensalter
unterschiedlich hoch ausfällt. Dies heißt auch, daß die
heiratsaufschiebende Wirkung des Berufszugangs sowie die
Differenzierung der Wartezeit durch Geschlechts- und
Kohortenzugehörigkeit altersbegrenzt ist.

In Tabelle 2 (siehe oben) sind die Anteile der bis zum Alter von
30 Noch-nicht-verheirateten und ihre mittlere Wartezeit bis zur
Heirat wiedergegeben. Wir sehen, daß die betreffenden sozialen

Gruppen sich in diesem Alter hinsichtlich ihrer
Heiratswahrscheinlichkeit bzw. ihrem mittleren Heiratsalter kaum
noch unterscheiden. Dagegen hat sich die soziale Differenzierung
darin niedergeschlagen, daß die Anteile von Personen, die erst
zwischen dem 30. und dem 40. Lebensjahr heiraten, unterschiedlich
groß sind. Es sind dies insbesondere Abiturienten, Akademiker,
Großstädter und Mitglieder der Kohorte 1929-31 sowie Personen
ohne berufliche Ausbildung. Es wird deutlich, daß der Aufschub
einer Familiengründung nicht unbedingt nachgeholt wird.
Allerdings hängt der Anteil derjenigen, die auf eine
Familiengründung "verzichten", davon ab, welches Alter man als
das kritische Lebensalter bezeichnet, nach dem eine
Familiengründung nicht mehr stattfindet.

2.2.2 Geschlechtsspezifische Bildungsdifferenzierung des Heiratsprozesses

Wir haben oben festgestellt, daß der Unterschied zwischen
Männern und Frauen hinsichtlich ihrer mittleren Dauer bis zur
Heirat nicht kleiner wird, wenn man ihre unterschiedlichen
Bildungswege berücksichtigt. Nun wollen wir prüfen, ob
Bildungsweg und Kohortenzugehörigkeit die Heiratschancen von
Männern stärker beeinflussen als die der Frauen. Zu diesem Zweck
schätzen wir getrennt die Effekte der Variablen für Männer und
Frauen.

Die Tabelle 5 zeigt, daß die Zusammensetzung der beiden
Geschlechtergruppen sich kaum unterscheidet. In den vorliegenden
drei Kohorten ist im Durchschnitt der Anteil der Abiturienten
höher als der der Abiturientinnen. Frauen beginnen zu einem
höheren Anteil als Männer ihren Berufsweg auf einem geringeren
Berufsstatus oder haben bis zur Ehe überhaupt nicht gearbeitet.
Ausbildungen in Fachschulen werden eher von Frauen als von
Männern absolviert. Der primäre Unterschied besteht jedoch im
beträchtlich größeren Anteil von Frauen, die keinerlei berufliche
Ausbildung besitzen. Dies betrifft vor allem Frauen der Kohorte
1929-31. In der Tabelle 6 (die entsprechenden Beta-Gewichte sind
in Tabelle A-2 wiedergegeben) sind die mittleren Wartezeiten von

Frauen und Männern in den einzelnen sozialen Gruppen
wiedergegeben, die durch Berufszugang, Ortstyp und
Kohortenzugehörigkeit bestimmt werden. Hier wurden nur die
mittleren Wartezeiten jener Gruppen wiedergegeben, die sich
signifikant unterscheiden.

Wir sehen, daß die mittlere Wartezeit bei Männern 10.7 und bei
Frauen 7.5 Jahre beträgt. Dieser Unterschied von drei Jahren wird
je nach sozialer Gruppe entweder reduziert oder verstärkt.
Folgende Altersverteilungen entsprechen diesem durchschnittlichen
Unterschied der Heiratswahrscheinlichkeit von Männern und Frauen
(siehe Schaubild 7; die entsprechenden Parameter sind in Tabelle
A-3). Die heiratsaufschiebende Wirkung eines höheren
Bildungsweges kommt bei Männern deutlicher zum Tragen als bei
Frauen. Insbesondere der Besuch eines Gymnasiums hat zur Folge,
daß die Familiengründung sehr weit auf ein späteres Lebensalter
verschoben wird. Im Vergleich zu Hauptschülern des eigenen
Geschlechts beträgt die mittlere Wartezeit bei Männern mit Abitur
durchschnittlich 7.6 Jahre, bei Frauen mit Abitur hingegen nur
2.8 Jahre mehr. Die zeitliche Spanne des Heiratsaufschubs ist bei
Abiturientinnen geringer als die Zeitspanne, die sie länger auf
die Schule gehen. Durch den Realschulbesuch wird die erste
Eheschließung bei den Männern um ca. zwei und bei den Frauen um
ca. ein Jahr aufgeschoben.

Wenn man den Zusammenhang von allgemeinbildender Ausbildung und
Heiratszeitpunkt betrachtet, dann bestätigt sich die Annahme, daß
die Familiengründung zugunsten instrumenteller Aktivitäten
zurückgestellt wird. Bei der beruflichen Ausbildung kehrt sich
diese Relation von Männern und Frauen um. Die akademische
Ausbildung, durch die die beruflichen Orientierungen erweitert
und die Chancen erhöht werden (zumindest in der hier betrachteten
historischen Periode) führt nicht zu einem Heiratsprozeß der
Familiengründung der Männer, der sich vom Heiratsprozeß von
Männern mit Facharbeiterausbildung unterscheidet. Männer mit
Lehre wie auch Männer mit einem universitären Abschluß heiraten
10 bis 11 Jahre nach Heiratsmündigkeit, d.h. im Alter von ca. 28
bis 29 Jahren. Frauen mit akademischer Ausbildung schieben
dagegen die Eheschließung auf. Sie entschließen sich

durchschnittlich drei Jahre später zur Heirat als ihre Geschlechtsgenossinnen mit einer Lehre. Dieser Effekt rührt jedoch nicht daher, daß Frauen mit Universitätsabschluß auch später als Männer mit Universitätsabschluß heiraten; bei Männern und Frauen mit akademischem Berufszugang wird die Heiratsentscheidung durchschnittlich 10 Jahre nach der Heiratsmündigkeit getroffen. Der "Universitätseffekt" ist bei den Frauen deshalb so ausgeprägt, weil Frauen mit einer Lehre drei Jahre früher als Männer mit Lehre heiraten (siehe auch Schaubild 8). Die Universitätsausbildung bedeutet in diesem Sinne keinen zusätzlichen Aufschub der Familiengründung. Die Tatsache, daß Frauen mit Lehre früher heiraten als Männer mit Lehre, führt dazu, daß der negative Einfluß der akademischen Bildung auf die Familiengründung bei Frauen überschätzt wird. Man könnte genauso gut sagen, daß nicht eine universitäre Ausbildung zum Aufschub der Heirat führt, sondern daß Frauen mit Lehre die Familiengründung vorziehen. Allerdings läßt sich nicht bestreiten, daß eine höhere berufliche Ausbildung den Heiratspozeß der Frauen an den der Männer angleicht.

In diesem Sinne wird die Geschlechtsdifferenzierung der Familiengründung durch eine anspruchsvolle Ausbildung nivelliert. Sehr deutlich kommt dies in Schaubild 9 zum Ausdruck. Männer ohne eine berufliche Ausbildung heiraten später als Frauen (siehe Schaubild 10). D.h. ein Berufseintritt ohne Qualifikation ist durchschnittlich mit einem Aufschub von ca. vier Jahren verbunden, wenn man dazu die Wartezeit der Männer mit Facharbeiterqualifikation vergleicht. Männer ohne berufliche Ausbildung heiraten damit durchschnittlich sogar später als Akademiker. Bei den Frauen spielte es in dieser historischen Periode (1950 bis 1974) für die Familiengründung keine Rolle, ob sie eine nicht-akademische Berufsausbildung absolvierten oder nicht.

Der soziale Status, mit dem der Berufsweg begonnen wird, hat nach den vorliegenden Schätzungen keine Bedeutung für den Zeitpunkt der Heirat. Dies kommt daher, daß er in hohem Maße mit dem Bildungsweg korreliert ist. Vorgezogene Eheschließungen sind mit einem fehlenden beruflichen Status vor der Heirat korreliert.

Dies gilt sowohl für Männer als auch für Frauen. Die
Berücksichtigung der Bildungseinflüsse kann zu einem gewissen
Teil die Vorverlagerung der Heirat in der Kohortenabfolge
erklären. Im Vergleich von Tabelle 7 und Tabelle 6 sehen wir, daß
die Kohortenunterschiede bei Männern und Frauen größer sind, wenn
man keine Bildungsunterschiede in Rechnung stellt. Allerdings
sind diese Bildungseffekte nicht so stark, daß sie die
Kohorteneffekte auf nicht-signifiaknte Unterschiede reduzieren
würden.

Bei den Männern ist kein eindeutiger Trend der
Kohortendifferenzierung festzustellen. Es gibt eine
Vorverlagerung um fast ein Jahr von Kohorte 1929-31 zu Kohorte
1939-41, die dann bei Kohorte 1949-51 durch Aufschub fast wieder
aufgeholt wird. Durch Berücksichtigung der Bildungseffekte werden
diese Kohortenunterschiede etwas reduziert.

Bei den Frauen verhält es sich umgekehrt. Durch die
Berücksichtigung der Bildungsdifferenzierung tritt ein deutlicher
Trend zur früheren Eheschließung noch stärker hervor. Er ist
zwischen der Kohorte 1929-31 und 1939-41 stärker ausgeprägt (der
Unterschied beträgt 1.6 Jahre) als zwischen der Kohorte 1939-41
und der Kohorte 1949-51 (0.8 Jahre). Auf den Aspekt der
geschlechtsspezifischen Kohortendifferenzierung kommen wir im
übernächsten Abschnitt noch zu sprechen. Im folgenden Kapitel
soll zunächst die globale Interkohortendifferenzierung der
Geschlechts- und Bildungsunterschiede beim Heiratsproß
untersucht werden.

2.2.3 Kohortenspezifische Einflüsse des Berufszugangs auf
      den Heiratsprozeß

Die Interkohortendifferenzierung des Bildungseinflusses auf den
Zeitpunkt der Eheschließung wird in den Tabellen 8 und 9
wiedergegeben. (Die entsprechenden Parameterschätzungen sind in
Tabelle A-4 und A-5 enthalten). Hauptsächlich hat sich der
Heiratsaufschub durch Besuch eines Gymnasiums auf die
Eheschließung verändert. Während bei der Kohorte 1929-31
Abiturienten 2.8 Jahre später als Hauptschüler heiraten, findet

der Übergang in die Ehe bei Abiturienten der Jahrgänge 1949-51
fast fünf Jahre später statt als bei Hauptschülern der gleichen
Jahrgänge. Dieser Unterschied läßt sich aber dadurch erklären,
daß die Hauptschüler ihre Entscheidung zur Familiengründung
vorgezogen haben, und nicht dadurch, daß die Abiturienten die
ihrige aufgeschoben haben. Dieser Wandel fand hauptsächlich
zwischen der Kohorte 1929-31 und der Kohorte 1939-41 statt. Im
Gegensatz dazu ist der Unterschied in den mittleren Wartezeiten
bis zur Heirat zwischen Hauptschülern und Realschülern zwischen
den Kohorten im wesentlichen stabil geblieben. Bei beiden Gruppen
ist eine Tendenz zur Vorverlagerung der Eheschließung zu
beobachten, die ca. zwei Jahre beträgt. In der Abfolge dieser
drei Kohorten bleibt die Verzögerung der Heirat durch eine
berufliche Ausbildung insgesamt stabil.

Man sollte jedoch die Vorverlagerung der Heirat bei Akademikern
und ihre wachsende Annäherung an das Heiratsalter der
Facharbeiter aufgrund der großen Streuung in beiden Gruppen nicht
überbewerten. Dies gilt auch für den Heiratsaufschub, der mit
einer fehlenden Ausbildung verbunden ist. Dieser ist als ein
historisch spezielles Phänomen der Kohorte 1929-31 zu betrachten.
Welche Gruppen zeigen also in der Kohortenabfolge eine Tendenz,
die Wartezeit bis zur Heirat zu verkürzen?

Im einzelnen sind dies die Frauen, die Hauptschüler, die
Realschüler, die Personen ohne Ausbildung. Die Vorverlagerung
der Wartezeit macht sich bei der Kohorte 1939-41 in allen
genannten Gruppen deutlicher bemerkbar als bei der Kohorte
1949-51. Die Differenz der Wartezeit von Männern und Frauen wird
von Kohorte zu Kohorte größer, sie verdoppelt sich von nahezu von
1.3 auf 2.3 Jahre. Es handelt sich dabei um die von den negativen
Bildungseffekten bereinigte Wartezeit. Bei der Kohorte 1949-51
ist der Geschlechterunterschied etwas größer, wenn die
Bildungseinflüsse einbezogen sind. Bei den anderen Kohorten
unterscheidet er sich dagegen nicht vom unbereinigten Wert.

2.2.4 Geschlechtsdifferenzierung des Heiratsprozesses bei den
      Kohorten 1929-31, 1939-41 und 1949-51

Im folgenden wollen wir die Geschlechtsdifferenzierung des
Heiratsprozesses in den einzelnen Kohorten genauer analysieren.
(Die geschätzten Parameter sind im Anhang in den Tabellen A-6 bis
A-9 wiedergegeben). Kohorte 1929-31 (siehe Tabelle 10): Für
diese Kohorte gilt die allgemeine Feststellung, daß ein höherer
Bildungsweg bei Männern und Frauen zum Aufschub der
Famliengründung führt, wobei dieser Effekt für Männer größer ist
als für Frauen. Abiturienten heiraten später als Hauptschüler und
auch später als Abiturientinnen. Der Unterschied liegt hierbei
weniger zwischen Hauptschülern und Abiturienten, sondern vielmehr
zwischen Abiturienten und Abiturientinnen. D.h in dieser Kohorte
wird die relativ geringe Geschlechtsdifferenzierung des
Heiratsprozesses durch Absolvierung eines gymnasialen
Bildungsweges verstärkt.

Eine akademische Ausbildung verschiebt den Heiratszeitpunkt
zusätzlich um durchschnittlich drei Jahre, sowohl bei Männern als
auch bei Frauen, wenn man sie mit ihren Geschlechtsgenossen
vergleicht, die eine Lehre absovierten. Die
Geschlechtsdifferenzierung in bezug auf die Wartezeit bis zur
Heirat wird durch den aufschiebenden Einfluß bei Personen mit
höherer Qualifikation nivelliert.

Bei Männern wird aber eine Eheschließung gleichermaßen durch
eine fehlende Ausbildung verschoben. Männer ohne Ausbildung
heiraten 3.5 Jahre später als Frauen ohne Ausbildung und drei
Jahre später als Männer mit Lehre. Für die Frauen bleibt der
Zeitpunkt der Familiengründung derselbe, unabhängig davon, ob sie
eine qualifizierte Ausbildung absolvieren oder ob sie auf eine
berufliche Qualifikation verzichten.

Kohorte 1939-41 (siehe Tabelle 11):

Bei der Kohorte 1939-41, deren Mitglieder zehn Jahre später
geboren wurden und deren Heiratsphase in die Zeit vom Ende der
Fünfziger bis zum Ende der Sechziger fiel, wird die
Geschlechtsdifferenzierung des Heiratsprozesses ausgeprägter. Im
Vergleich zu Frauen der Kohorte 1929-31 heiraten Frauen mit
Abitur in dieser Kohorte um denselben Zeitraum später als
Hauptschüler. Bei Männern wächst die Differenz der mittleren
Wartezeiten von Hauptschülern und Abiturienten und beträgt in
dieser Kohorte fast sieben Jahre.

Das gleiche Muster tritt beim geschlechtsspezifischen Einfluß der
beruflichen Ausbildung auf. Während bei den Frauen der
Heiratsaufschub durch eine akademische Ausbildung um drei Jahre
als der gleiche erhalten bleibt wie bei den Frauen der
vorhergehenden Kohorte, heiraten Männer mit Universitätsabschluß
im Mittel um ein Jahr früher als Facharbeiter. Insgesamt bedeutet
dies, daß sich die Heiratsprozesse von Frauen und Männern mit
akademischer Ausbildung noch stärker angleichen als in der
Kohorte 1929-31. Trotz prosperierender Wirtschaft stellt eine
fehlende Ausbildung bei den Männern dieser Jahrgänge ein größeres
Handicap für die Familiengründung dar. Sie müssen länger bis zur
Eheschließung warten als vergleichbare Männer der vorhergehenden
Kohorten und erheblich länger als Frauen. Für Frauen dagegen
bedeutet eine fehlende Berufsausbildung sogar eine Verkürzung der
Wartezeit bis zur Heirat.

Kohorte 1949-51 (siehe Tabelle 12):

Die Männer und Frauen in der jüngsten der drei Kohorten wachsen
in den fünziger und sechziger Jahren auf, und ihre Heiratsphase
fällt in die Zeit vom Ende der sechziger bis zum Ende der
siebziger Jahre. Die Geschlechtssdifferenzierung der
Bildungseinflüsse bleiben auf dem gleichen hohen Niveau der
Vorkohorte bestehen. Im Gegensatz zur Kohorte 1929-31 muß man
aber feststellen, daß eine akademische Ausbildung bei Männern der

jüngsten Kohorte nicht mit einem zusätzlichen Aufschub der Heirat
verbunden ist. Die Wartezeit von Frauen mit Universitätsabschluß
ist bei allen drei Kohorten länger als die Wartezeit von Frauen
mit Lehre. Insgesamt bleibt die Angleichung des Heiratsalters
von Frauen und Männern mit akademischer Bildung erhalten.
Dahinter verbirgt sich jedoch eine Geschlechtsdifferenzierung,
die erst durch Universitätsausbildung bei den Frauen aufgehoben
wird. Für Männer in der Kohorte 1949-51 bedeutet eine fehlende
Ausbildung einen besonders langen Aufschub der Heirat. Sie
heiraten durchschnittlich mehr als 10 Jahre später als Frauen
ohne berufliche Ausbildung und 8.4 Jahre später als Männer, die
eine Ausbildung haben.

Insgesamt zeichnet sich die jüngste Kohorte durch eine besonders
starke Geschlechtsdifferenzierung im Prozeß der Familiengründung
aus. Sie rührt daher, daß die Kohortendifferenzierung bei Frauen
und Männern unterschiedliche Schwerpunkte hat. Bei den Frauen
(vgl. Tabelle 13) ist und bleibt sowohl die allgemeine wie auch
die berufliche Bildungsinvestition das entscheidende
differenzierende Kriterium für die Wahl des Zeitpunktes zur
Familiengründung. Für die Eheschließung von Frauen ist es
unerheblich, ob sie einen Beruf erlernt haben oder nicht.

Im Fall der Männer (vgl. Tabelle 14) gilt dies nur für die
allgemeine Bildungsinvestition. Diese ist mit einem Aufschub der
Heirat verbunden, wobei dieser Effekt von Kohorte zu Kohorte
stärker wird. Eine hohe berufliche Qualifikation führt nicht
dazu, daß Männer den Zeitpunkt der Heirat aufschieben. Hingegen
kristallisiert sich in der Abfolge der drei ausgewählten Kohorten
eine fehlende Qualifikation immer stärker als das zentrale
Merkmal heraus, mit dem besonders niedrige Heiratschancen
assoziiert sind. An dieser Stelle kommt die
Geschlechtstypisierung der Heiratschancen besonders deutlich zum
Ausdruck.

2.2.5 Dekomposition der Kohorten- und Geschlechtsdifferenzierung
       des Heiratsprozesses: Verhaltens-, Kompositions- und
       Alterseffekte
In diesem Abschnitt soll geprüft werden, aus welchen Komponenten

sich die beobachteten Differenzen der mittleren Wartezeiten bis
zur Eheschließung zwischen Männern und Frauen und zwischen den
Kohorten zusammensetzen. Mit der von uns durchgeführten
Modellschätzung kann man folgendes Gedankenexperiment empirisch
umsetzen: Wie würde die mittlere Dauer bis zur Heirat bei Frauen
aussehen, wenn sie die gleiche Bildungsbeteiligung wie Männer
hätten?

Hierbei will ich zwei Aspekte der Bildungsbeteiligung
unterscheiden: - **Bildungseffekt**: damit ist der Stellenwert von
Bildungsmerkmalen für den Zeitpunkt der Heirat gemeint, wie er in
den Abschnitten 2.2.1 bis 2.2.4 beschrieben wurde. **Effekt der
Bildungsverteilung**: damit soll der Anteil der einzelnen
Bildungsgruppen in einer Population bezeichnet werden.

In den Tabellen A-10 (Frauen) und A-11 (Männer) können wir
sehen, daß sich das Qualifikationsprofil von Männern und Frauen
verbessert und zunehmend angeglichen hat. Ich möchte an dieser
Stelle nicht näher darauf eingehen; eine detaillierte Diskussion
gibt BLOSSFELD (1985). Außerdem kann man die Frage aufwerfen,
wie sich der Heiratsprozeß von Frauen ändern würde, wenn bei
ihnen der gleiche Alterseffekt anzutreffen wäre wie bei Männern.
Tabelle 15 zeigt das Ergebnis, das man erhält, wenn man die
Gruppenmerkmale der Frauen an die der Männer für die drei
ausgewählten Kohorten angleicht. Zunächst sei die erste
Simulation betrachtet. Frauen unterscheiden sich hier von den
Männern nur hinsichtlich der Bildungseffekte. In diesem Fall
stimmt die mittlere Wartezeit von Frauen und Männern überein, der
Bildungseffekt trägt insgesamt nur wenig zur
Geschlechtsdifferenzierung des Heiratsprozesses bei.

Bei der jüngsten Kohorte, in der die Bildungseffekte stärker
geworden sind, haben sie einen starken differentiellen Effekt in
bezug auf das Heiratsverhalten von Männern und Frauen. Der
Unterschied zwischen Männern und Frauen würde am deutlichsten
nivelliert sein, wenn sich die Geschlechtergruppen nur in der
Bildungsverteilung unterscheiden würden. Anders ausgedrückt, wenn
die Altersabhängigkeit der Heiratswahrscheinlichkeit und der

familiale Stellenwert des Bildungsweges bei Frauen und Männer, gleich wären, würden Frauen sogar später als Männer heiraten.

In der ältesten Kohorte würde die Angleichung des Alterseffektes von Frauen an den der Männer dazu führen, daß ihre Wartezeit um 0.6 Jahre sinken würde, die Angleichung des Bildungseinflusses würde zu einer Verminderung der mittleren Wartezeit um um 0.5 Jahre führen. Würden dagegen der Altersprozeß der Männer und deren Bildungseffekte für die Frauen gelten, dann würden sie später als Männer heiraten.

Bei der mittleren Kohorte hätte die Angleichung der Bildungsstruktur noch größere Effekte. Sowohl die Angleichung des Alterseffektes wie auch die des Bildungseinflusses hat zur Folge, daß der mittlere Unterschied der Wartezeiten von Männern und Frauen fast verschwindet. Man kann sagen, daß Frauen später heiraten würden, wenn sie bei gleichem Altersprozeß auch die gleichen Bildungseffekte hätten wie die Männer.

Dies gilt gleichermaßen auch für die Kohorte 1949-51: erst wenn man bei Frauen die gleichen Altersprozesse und Bildungsverteilungen wie bei Männern finden würde, gäbe es keinen geschlechtsspezifischen Heiratsprozeß in dieser Kohorte. Nur die Angleichung des Bildungseinflußes reduziert zwar den mittleren Altersunterschied, wie das auch für die Angleichung nur des Altersprozesses zutrifft. Keines der beiden Merkmale allein kann aber die Heiratsprozesse angleichen.

In gleicher Weise wollen wir nun die Angleichungssimulation für die Bildungsmerkmale der Kohorten durchführen. Wir gleichen die Bildungsmerkmale der Kohorten 1939-41 und 1949-51 an die Bildungsmerkmale der Kohorte 1929-31 an.

Betrachten wir zunächst die Ergebnisse, die sich für die Frauen ergaben (siehe Tabelle 16). Wenn sich zwischen den Kohorten der Frauen nur der Alterseffekt verändert hätte, dann hätte eine Vorverlagerung der Heirat deutlicher sein müssen, als es tatsächlich der Fall war. D.h. die Veränderung der Bildungsstruktur in den Kohorten 1939-41 und 1949-51 hat der

Tendenz zur Vorverlagerung der Heirat entgegengearbeitet. Wäre andererseits der Altersprozeß der gleiche wie bei der Kohorte 1929-31 geblieben, dann wäre die Bildungsexpansion mit einem durchschnittlichen Aufschub der Heirat um ein Jahr verbunden gewesen.

Man kann also schlußfolgern, daß bei den Frauen eine von den Bildungsinvestitionen unabhängige Vorverlagerung der Heirat stattgefunden hat. Diese wurde aber durch die Bildungsexpansion, insbesondere durch den veränderten Stellenwert der Qualifikation für die Familiengründung in ihrer Bedeutung gemindert. Der heiratsaufschiebende Effekt der Bildunsexpansion wurde durch eine parallele Entwicklung, die zur Vorverlagerung der Heiratsentscheidung führte, konterkariert. Auch bei den Männern hätte es eine solche Vorverlagerung der Familiengründung gegeben, und zwar von 10 auf 8.5 Jahre bei der jüngsten Kohorte (vgl. Tabelle 17). Insbesondere bei der Kohorte 1949-51 wurde aber diese Entwicklung durch die Erweiterung heiratsaufschiebender Bildungswege gebremst. Dies gilt vor allem für die Männer der Kohorte 1949-51. Wie bei den Frauen waren dafür weniger die expandierten Anteile höherer Bildungsgruppen verantwortlich, sondern hauptsächlich die verstärkten Bildungseinflüsse. Die stärkere Verwertbarkeit von Bildungsinvestitionen hätte auch bei den Männern zum stärkeren Aufschub der Familienbildung geführt, wenn es nicht eine davon unabhängige Tendenz zur Vorverlagerung der Eheschließung gegeben hätte.

2.2.6 Altersabhängigkeit von Heirat und Berufszugang

Wie wird nun die Altersabhängigkeit der Heiratsentscheidung durch die Effekte unterschiedlicher Berufszugänge beeinflußt?

Schaubild 11, dem die Parameter in Tabelle 1 entsprechen, zeigt, daß die Altersabhängigkeit durch die Eweiterung der Modellkovariaten stärker wird. Wir wollen uns den Vorgang noch einmal vergegenwärtigen: Zunächst gründen wir unsere Voraussage der Heiratswahrscheinlichkeit nur auf das Alter. Wir sind davon ausgegangen, daß die Heiratsrate mit steigendem Alter bis zu einem Punkt tm ansteigt und nach dessen Überschreiten wieder

fällt. Unser Modell nimmt als zusätzliche Information die
Merkmale des Berufszugangs zur Vorhersage der Heiratsrate hinzu.
Aufgrund der obigen Analyse können wir davon ausgehen, daß ein
höherer Berufszugang die Heiratswahrscheinlichkeit zu jedem
Zeitpunkt vermindert. Folgende Simulation bringt dies deutlich
zum Ausdruck. Wenn wir zur Dauerabhängigkeit des Basismodells
(dem Modell ohne Kovariaten) den gesamten Einfluß der Kovariaten
des Vollmodells hinnehmen, dann wird die nicht-monotone
Verteilung der Hazardrate stark abgeschwächt (siehe Schaubild
12). Wie kann man auf der Basis dieser Zusammenhänge die
bereinigte Altersabhängigkeit im Vollmodell interpretieren?

Zunächst können wir bei allen Subgruppen, den Männern und Frauen
der drei Kohorten, festhalten, daß der Einfluß der Wartezeit auf
die Heiratsrate, auch unter Berücksichtigung der
Bildungsvariablen, weiterhin hoch ist und signifikant bleibt
(vgl. die Tabellen A-2, A-5, A-7 und A-8). Die
Bildungsheterogenität der Subgruppen reicht demnach hier nicht
zur Erklärung aus.

Wenn wir diesen Alterseffekt mit jenem Alterseffekt
vergleichen, der sich ergibt, wenn man keine Bildungsvariablen
berücksichtigt, dann kann man feststellen, daß er bei allen
Subgruppen größer ist. Wenn man keine Kovariaten zur Prädiktion
der Heiratsrate hinzunimmt, dann kann man sagen, daß sich im
Alterseffekt die gesamte beobachtbare Heterogenität der
Populationen hinsichtlich der Heiratswahrscheinlichkeit
ausdrückt. Berücksichtigt man Kovariaten, dann stellt man die
durch sie evozierte Heterogenität in Rechnung. Der Alterseffekt
spiegelt dann weniger Heterogenität wider und müßte kleiner
werden.

Dies entspricht dem üblichen Heterogenitätsargument. Man kann
aber auch genau umgekehrt argumentieren: Die maximale
beobachtbare Heterogenität der Heiratswahrscheinlichkeit in einer
Population entspricht einer alterskonstanten Rate. Eine konstante
Rate bedeutet, daß der Zeitpunkt der Heirat für die
Wahrscheinlichkeit ohne Bedeutung ist, d.h. daß die
Heiratswahrscheinlichkeit in jeder Altersstufe gleich ist. Dies

wiederum heißt, daß die Heiratsentscheidungen über alle
Zeitpunkte gleichmäßig verteilt sind. In dem Maße, wie wir durch
exogene Variablen diese Streuung der Heiratswahrscheinlichkeit
reduzieren, muß dementsprechend die Altersbeliebigkeit reduziert
und damit eine größere Altersabhängigkeit beobachtbar sein. Wenn
wir durch Konstanthaltung diese Einflüsse ausschließen, müßte die
Altersabhängigkeit und das Niveau der Heiratsrate größer werden.

Betrachten wir nun unter diesen Gesichtspunkten das Verhältnis
von Basismodell und Vollmodell in den einzelnen Subgruppen. Bei
den Männern bedeutet der Berufszugang in allen Kohorten, daß es
zu einer Nivellierung der Altersabhängigkeit der
Heiratswahrscheinlichkeit kommt. D.h. die Differenzierung einer
Population hinsichtlich ihrer Bildungswege erzeugt nicht eine
Altersabhängigkeit, sondern einen entgegengesetzten Effekt, indem
sie generell mit einem Aufschub der Heirat bzw. einer Reduktion
der Heiratswahrscheinlichkeit einhergeht.

Allerdings ist diese "Einflußkraft" des Berufszugangs auf die
Heiratswahrscheinlichkeit insgesamt nicht besonders groß. Bei der
Kohorte 1929-31 (siehe Schaubild 13) ist sie am geringsten. Auch
bei der Kohorte 1939-41 (siehe Schaubild 14) spielt sie kaum eine
größere Rolle. Bei der Kohorte 1949-51 führt die Bereinigung der
Bildungseffekte am stärksten zu einer Vergrößerung der
Altersabhängigkeit (Schaubild 15). Man muß dabei
berücksichtigen, daß die Männer in der Kohorte 1949-51 sich
dadurch auszeichnen, daß bei ihnen die geringste
Alterskonzentration im Basismodell zutage trat.

Wenn man den Bildungseffekt einbezieht, so nähert sich die
Altersverteilung der Heiratsrate bei den Jahrgängen 1949-51 an
die Altersverteilung der Jahrgänge 1929-31 an (Schaubild 16).
Die Kohorte 1939-41 unterscheidet sich von den anderen darin, daß
sie die größte Alterskonzentration und die höchsten Heiratsraten
aufweist. Beides wird durch Bildungseinflüsse kaum verändert.
Dieser Zusammenhang von Bildungseffekt und Altersgradierung gilt
gleichermaßen für den Heiratsprozeß von Frauen. In einem
nennenswertem Umfang finden wir die Reduktion der
Altersgradierung durch den Bildungseinfluß nur bei Frauen der

Kohorten 1939-41 und 1949-51 (vgl. Schaubild 18 und 19). Bei der
ältesten Kohorte nehmen die Bildungseinflüsse kaum Einfluß auf
die Altersabhängigkeit (vgl. Schaubild 17).

Wenn man die drei Kohorten jeweils bei Frauen und Männern
hinsichtlich der Altersverteilungen miteinander vergleicht,
zeichnet sich folgendes Bild ab: a) Faktisch gibt es eine
Vorverlagerung des Heiratsprozesses, (dies zeigte sich schon oben
anhand der mittleren Wartezeit). b) Die Altersgradierung der
Heiratsentscheidung wird durch diese Bildungseinflüsse gedämpft.
Wie verändert sich die Altersabhängigkeit in der Kohortenabfolge,
wenn man die bremsenden Einflüsse des Berufszuganges und seines
sozialen Wandels in Rechnung stellt, indem man sie gewissermaßen
"abzieht"?. Da hier nicht mehr die Vorverlagerung oder der
Aufschub der Heiratsentscheidung per se im Vordergrund steht,
wollen wir die Heiratsprozesse der drei Kohorten (für Männer und
Frauen getrennt) zeitlich standardisieren. Dies erreichen wir
dadurch, daß wir für alle drei Kohorten die maximale
Heiratswahrscheinlichkeit auf den gleichen Zeitpunkt setzen.
Dafür wählen wir den Zeitpunkt, zu dem die älteste Kohorte die
maximale Heiratswahrscheinlichkeit aufweist. Die so
standardisierten Heiratsprozesse der drei Kohorten stehen bei den
Männern in folgendem Verhältnis zueinander (siehe Schaubilder
20,21).

Wenn man den unterschiedlichen Stellenwert des Berufszugangs bei
den drei Kohorten in Betracht zieht, wird deutlich, daß die
Alterskonzentration der Heiratsentscheidung zwischen der Kohorte
1929-31 und der Kohorte 1949-51 abnimmt. Der Unterschied ist
allerdings nicht sehr groß. Bei den Männern der Jahrgänge
1939-41 findet man dagegen eine Zunahme der Altersabhängigkeit,
sowohl relativ zu der Kohorte 1929-31 wie auch zur Kohorte
1949-51. Bei den Frauen (vgl. Schaubild 22 und 23) sind noch
deutlichere Veränderungen der bereinigten Altersabhängigkeit zu
beobachten.Sie betreffen vor allem den Zeitraum zwischen der
Kohorte 1939-41 und 1929-31. Die Ähnlichkeit der Heiratsprozesse
zwischen den Kohorten 1939-41 und 1949-51 ist sehr groß, aber nur
wenn man die Vorverlagerung und die Bildungseffekte der Kohorte
1949-51 außer acht läßt.

3. Zusammenfassung und Schlußfolgerungen

Der Berufszugang hat für den Heiratsprozeß zwei Bedeutungen: 1.
Die Dauer der formalen schulischen Bildung nimmt Einfluß darauf,
zu welchem Zeitpunkt im individuellen Lebensverlauf die
Heiratsentscheidung getroffen wird. 2. Die mit unterschiedlichen
Optionen verknüpften beruflichen Qualifikationen tragen zu einer
weiteren Differenzierung des Heiratsprozesses bei. Die
heiratsaufschiebende Wirkung weiterführender Bildungsaktivitäten
gilt gleichermaßen für Männer und für Frauen. Auf allen
Bildungsniveaus heiraten Frauen früher als Männer. Allerdings
fällt auf, daß es für Frauen eine andere Bedeutung hat als für
Männer, wenn sie das Gymnasium absolvieren. Während Abiturienten
ihre Heirat über 5 Jahre mehr als Realschulabsolventen und fast
sieben Jahre mehr als Hauptschüler aufschieben, entspricht bei
Abiturientinnen die Heiratsverzögerung, wie sie durch ihren
Erwerb höherer Bildung entsteht, nicht ihrer längeren Schuldauer.

Die geschlechtsspezifischen Wartezeitunterschiede zwischen
Hauptschülern und Realschülern kann man nicht als Indikatoren der
Orientierung von Männern und Frauen hinsichtlich der
Familienbildung interpretieren. Denn es konnte gezeigt werden,
daß es mit der Ausweitung der instrumentellen Aktivitäten durch
Realschulbildung sowohl bei Männern als auch bei Frauen zum
Aufschub der Familiengründung kommt.

Wie läßt es sich aber verstehen, daß es gerade im Falle der
gymnasialen Bildungsbeteiligung zur Aktualisierung von
Geschlechtsrollen in dem Sinne kommt, daß Männer die
Familienbildung länger aufschieben als Frauen?

Wir können für Frauen in der untersuchten historischen Periode
die Lebensphase, in der für sie eine Familiengründung möglich
erscheint, bis Ende Zwanzig datieren. Das wahrscheinlichste
Heiratsalter fällt in die Mitte Zwanzig.

Diesem "kritischen" Lebensalter sind aber Abiturientinnen

infolge ihrer langen Moratoriumszeit so nahe gekommen, daß ihnen
nicht mehr viel Zeit für andere extensive nicht-familiale
Aktivitäten übrigbleibt. Für Männer hingegen spielt dieses
"kritische" Lebensalter eine viel geringere Rolle bzw. es ist
weit in ein höheres Lebensalter verschoben, in der die
Stabilisierung der psychisch-sozialen Bedürfnisbefriedigung für
sie dringlicher wird. D.h. Männer können sich mehr Lebenszeit für
andere Dinge nehmen, bevor sie die psychisch-ökonomische Bindung
einer Familie eingehen. Dies gilt besonders für jene Männer, die
sich aufgrund ihrer höheren allgemeinen Schulbildung
diversifizierte Chancen und Bedürfnisse geschaffen haben. Die
Erklärung durch die ökonomische Stabilisierung führt hier nicht
weiter, da sie unzureichend die Tatsache erklärt, daß es gerade
Abiturienten sind, die Eheschließung aufschieben. Da aber
Abiturienten mehr Karrierechancen mit höheren Einkommenszielen
haben, werden sie die Eheschließung als Familiengründung länger
aufschieben, um zunächst berufliche Ziele zu erreichen. Erst zu
einem späteren Zeitpunkt in ihrem Leben ist es ihnen möglich,
ihre berufliche Position zu etablieren.

Die Geschlechtsspezifizität des Heiratsaufschubs aufgrund der
weiterführenden Schulbildung ist ein Ergebnis des sozialen
Wandels von 1950 bis 1980, wie er vor allem bei der Kohorte
1949-51 zum Ausdruck kommt. Bei allen drei Kohorten bringt die
höhere Bildung eine Verzögerung der Eheschließung und
Familiengründung mit sich. Aber erst in Kohorte 1949-51 gewinnt
die höhere Schulbildung bei Männern eine solche Bedeutung, daß es
dazu kommt, daß die Familienbildung deutlich zurückgestellt wird.

Allerdings handelt es sich dabei nicht um einen linearen Trend.
Das Abitur hat schon für die Männer der Kohorte 1939-41 die
gleiche negative Verzögerung der Familienbildung zur Folge. Bei
den Frauen findet sich die geringere Bedeutung des Abiturs für
die Familienbildung sowohl bei der Kohorte 1929-31 als auch bei
der Kohorte 1949-51. Abiturientinnen jedoch, die zwischen 1939-41
geboren wurden und die ihre "kritische" Heiratsphase in den
sechziger Jahren durchliefen, zeigen ein davon abweichendes
Muster. Diese Frauen warten länger bis zur Heirat als jene
Frauen, die zehn Jahre vor oder nach ihnen geboren wurden.

Es ist schwierig, die vorliegende Kohortendifferenzierung bei
Männern und Frauen auf die gleiche Weise zu interpretieren.
Zunächst steht die Frage im Raum, warum die Frauen der Jahrgänge
1939-41 ihre höhere Bildung stärker für außerfamiliale
Aktivitäten nutzten als die Frauen der jüngsten Jahrgänge. Eine
befriedigende Antwort kann angesichts der Komplexität der
historischen Situation nicht gegeben werden. Es fällt jedoch auf,
daß gerade der Schulabgang dieser Abiturientinnen-Jahrgänge mit
einer prosperierenden Wirtschaft zusammenfiel. Es war die
eigentliche Phase der "Goldenen Fünfziger Jahre", nämlich die
Periode zwischen 1961 und 1967. Für Frauen, die zehn Jahre
später ihr Abitur machten, fällt die aufgeschobene "kritische"
Heiratsphase in eine Zeit, die durch die ersten Ausläufer der
ökonomischen Rezession - "Ölkrise" 1973/74 - gekennzeichnet ist.
Die Frauen hingegen, die in der Nachkriegszeit eine höhere
Schulbildung absolvierten, hatten weder ökonomische noch soziale
Chancen, zunächst auf eine Familiengründung zu verzichten.
Denzufolge bilden sie jene Gruppe, für die
Bildungsdifferenzierung eine nur geringe Differenzierung des
Heiratsalters bedeutete. Hauptschülerinnen heirateten in dieser
Kohorte nicht sehr viel früher als Abiturientinnen.

Neben den schlechten sozio-ökonomischen Bedingungen und der
traditionellen Rollenorientierung führte das erzwungene
Partnerwahl-Moratorium der unmittelbaren Nachkriegszeit zur
Verspätung der Familiengründung. Die Kohorte 1939-41 erlebte
einen wieder normalisierten Heirats- und Arbeitsmarkt. Der Vor-
und Nachkriegsbabyboom führte dazu, daß den Frauen dieser
Jahrgänge mehr ältere Ehepartner gegenüberstanden. Zugleich
hatten diese Männer sehr viel günstigere Arbeitsmarkt- und
Einkommenschancen als die heiratsfähigen Männer zehn Jahre zuvor.
Die Frauen der Kohorte 1949-51 konnten zwar auch früher heiraten
und eine Familie gründen, aber die Tendenz der Vorverlagerung ist
bei ihnen schon deutlich gedämpft.

In bezug auf die Kohortendifferenzierung der Männer können wir
ebenfalls eine Vorverlagerung der Familiengründung als Ausdruck
verbesserter ökonomischer Chancen festzustellen. Allerdings

tritt diese Tendenz nur bei Hauptschülern und Realschülern als
lineare Entwicklung auf. Bei Abiturienten interferiert sie
hingegen mit der sehr deutlichen Verlängerung der Wartezeit bis
zur Heirat bei den Kohorten 1939-41 und 1949-51. Diese
gegenläufigen Effekte kompensieren sich bei den Männern in der
Gesamtpopulation derart, daß insgesamt nur eine geringe
Verkürzung der Wartezeit bis zur Familiengründung zu beobachten
ist. Hier sei zunächst nur festzuhalten, daß die Vorverlagerung
auch bei den Männern mit Abitur zu finden ist, wobei dies ab
Kohorte 1939-41 auf einem höheren Altersniveau gilt. Es ist
hervorzuheben, daß der allgemeinbildende Teil des Berufszuganges
sich in Aufschub der Heirat niederschlägt. Obwohl diese Bedeutung
geschlechtsspezifisch ist, konnte die Geschlechtsdifferenzierung
der Familiengründung nur zu einem geringen Teil auf die
unterschiedliche Relevanz der allgemeinen Qualifizierung für
Männer und Frauen zurückgeführt werden. Auch wenn man die
unterschiedliche Bedeutung des Schulabschlusses für Männer und
Frauen berücksichtigt, bleibt der Altersunterschied zwischen
Männern und Frauen im wesentlichen bestehen. Auch der veränderte
Stellenwert der höheren Bildung reicht nicht aus, die
Vorverlagerung der Heirat zu erklären. Man kann aber zeigen, daß
die wachsende Bedeutung der gymnasialen Bildung für die Männern
zur Kompensation des Vorverlagerungstrends führte. Die
Gegenläufigkeit dieser Trends ist auch bei den Frauen zu finden,
aber nur bei der Kohorte 1939-41, die unter günstigen
ökonomischen und sozialen Bedingungen aufgewachsen ist. Die
Bedeutung der beruflichen Ausbildung für die Wahl des
Heiratszeitpunkts besteht in ähnlicher Weise im Aufschub der
Heirat.

Allerdings kann man nicht davon ausgehen, daß
Bildungsinvestition und Heiratsalter in einer linearen Beziehung
zueinander stehen, sondern vielmehr besteht eine U-förmige
Beziehung zwischen dem Ausmaß der Qualifikation und der Wartezeit
bis zur Heirat. Die Ursache dafür ist darin zu sehen, daß die
fehlende Qualifikation nicht als eine zeitliche Entlastung
zugunsten der Familiengründung zu betrachten ist. Eine fehlende
Qualifikation ist mit einer längereren Wartezeit bis zur Heirat
verbunden. Dies gilt jedoch nur für Männer, deren Heiratschancen

im wesentlichen über ihre Fähigkeiten zur materiellen Sicherung
des Lebensunterhalts in der Familie bestimmt sind.
Dementsprechend heiraten Akademiker nicht sehr viel später als
Facharbeiter. Die verlängerte Moratoriumsphase wird quasi durch
die besseren Erwerbschancen kompensiert.

Kann es aber auch sein, daß bei ihnen der Aufschub durch die
Bildungsaktivitäten im Hinblick restriktiver Altersnormen durch
beschleunigte Heirat nach Ausbildungsende wettgemacht wird?

Mit unseren Ergebnissen können wir nicht entscheiden, ob
Altersnormen oder ökonomische Bedingungen zur Kompensation des
Aufschubs führen. In bezug auf die Entscheidung spielt es für
Frauen keine Rolle, welche Ausbildung sie haben. Unabhängig von
ihrem Bildungsniveau heiraten sie im Durchschnitt sieben Jahre
nach ihrer Heiratsmündigkeit, d.h. im Alter von 23 Lebensjahren.
Ihre Entscheidung zur Heirat wird wesentlich davon beeinflußt, ob
sie eine akademische Ausbildung absolvieren. In diesem Falle
entsteht der Eindruck, daß Frauen mit den besseren ökonomischen
und sozialen Chancen, die sie aufgrund ihrer akademischen Bildung
erworben haben, eher auf die Familiengründung verzichten. Dies
ist jedoch nur teilweise richtig. Im Vergleich zu Männern
bedeutet für Frauen ein Studium ein zusätzliches familiales
Moratorium. Dies kann dadurch verursacht sein, daß Frauen mit
hoher Qualifikation geringere Heiratschancen haben, weil sie
nicht die traditionelle Frauenrolle übernehmen wollen bzw. können
bzw. diese zurückstellen, um ihre Berufschancen zu realisieren.

Der Vergleich mit der Wartezeit der akademisch gebildeten Männer
spricht aber gegen diese Interpretation. Denn akademisch
gebildete Frauen schieben ihre Familiengründung nicht länger auf
als Männer mit Hochschulausbildung. Im Gegenteil, sie heiraten im
Durchschnitt im Alter von 26 Jahren und die Männer im
Durchschnitt von knapp 27 Jahren. Indem die akademische
Qualifikation sich bei Frauen in die kritische Lebensphase der
Familiengründung schiebt, führt sie sie tendenziell vor die
Entscheidung, entweder eine Familie zu gründen oder auf eine
Familie zu verzichten. Diese Beziehung zwischen kritischer
Lebensphase der Familiengründung und beruflichem Bildungsweg hat

sich bei Frauen in der Zeit von 1950 bis 1980 stärker
herausgebildet, obwohl oder gerade weil die Wartezeit bis zur
Familiengründung bei Frauen immer kürzer geworden ist. Während
bei den Frauen der Kohorte 1929-31 aufgrund der traditionellen
Familienorientierung und schlechter Heiratschancen die
akademische Ausbildung kaum zum Aufschub der Familiengründung
führte, hat sie bei den beiden anderen Kohorten eine zusätzliche
familiale Moratoriumszeit bedeutet. Dieser Eindruck entsteht
jedoch dadurch, daß sich Frauen mit Hochschulausbildung stärker
von Frauen mit Lehre hinsichtlich ihrer mittleren Wartezeit bis
zur Heirat unterscheiden. Wobei sich die mittlere Wartezeit der
Frauen mit akademischer Bildung nur unwesentlich verändert hat.

Dies spricht in zweifacher Hinsicht für die Signifikanz der
kritischen Heiratsphase im Lebensverlauf einer Frau.

Zum einen müssen Frauen mit hoher Bildung die zeitlich begrenzte
Phase einer möglichen Familiengründung bei ihren familialen und
beruflichen Entscheidungen in Rechnung stellen. Trotz des
sozialen Wandels wurde dieser kritische Zeitpunkt - er liegt bei
allen Frauen ungefähr beim 27. Lebensjahr - nicht wesentlich
verschoben.

Zum anderen können jene Frauen vom ökonomischen und sozialen
Wandel dieser Zeit profitieren, die aufgrund ihres
Ausbildungsweges hinsichtlich des Zeitpunktes ihrer
Familiengründung weniger festgelegt sind. Von Kohorte zu Kohorte
verkürzen Frauen ohne berufliche Ausbildung oder Frauen mit einer
Lehre ihre Wartezeit bis zur Familiengründung. Vor allem der
Unterschied zwischen der Kohorte 1929-31 und der Kohorte 1939-41
reflektiert die erhöhten Heiratschancen der Frauen der Kohorte
1939-41, die in den "Goldenen Fünzigern und Sechzigern" in einem
gesellschaftlichen Klima der familialen Restauration mehr Männer
mit besserer Zukunft fanden.

Eine weitere Verkürzung der Wartezeit bei der Kohorte 1949-51 ist
nicht so ausgeprägt wie bei der Kohorte 1939-41. Die meisten
dieser Frauen heiraten in den siebziger Jahren, und ihre Chancen
zur Familiengründung sind im Heiratspool und was die ökonomische

Lage der potentiellen Ehemänner betrifft, besser als je zuvor. Es hat sich aber mittlerweile eine gesellschaftliche Wandlung vollzogen, die die Vorverlagerung der Familiengründung dämpft. Wenn der Stellenwert der beruflichen und allgemeinen Bildung gleich geblieben wäre, dann wäre die Verkürzung der Wartezeit bis zur Familiengründung noch stärker ausgefallen. Hier deutet sich der Niederschlag der Bildungsexpansion auf die familiale Lebensorganisation von Frauen an. Er macht sich aber noch nicht im Aufschub bzw. Verzicht bemerkbar, weil die Frauen der Kohorte 1949-51 mit ihrer durchschnittlich frühen Heiratsphase nur in die Anfangsphase der Bildungsexpansion gekommen sind.

Bei Männern, deren Heiratsphase insgesamt später im Lebensverlauf liegt, werden die Effekte der Bildungsexpansion tatsächlich als neuer Aufschub der Familiengründung sichtbar, nach dem zuvor eine Verkürzung der Wartezeit stattgefunden hatte. Zugleich hat sich im Zuge dieses sozialen Wandels auch die geschlechtsrollenspezifische Fixierung der Männer verstärkt. Männer ohne Qualifikation haben in einer qualifikationsorientierten Gesellschaft die längsten Wartezeiten bis zur Familiengründung. Dafür sind ihre schlechten Chancen verantwortlich, die sie weniger im Arbeitsmarkt als im Heiratsmarkt haben. Man sieht dies daran, daß vom Wohlstand und Zukunftsoptimismus der fünziger und sechziger Jahre, wie sie die Kohorte 1939-41 erlebte, nur die Männer mit akademischem Abschluß für ihre Familiengründung profitieren konnten. Und umgekehrt, von den beginnenden schlechten Zeiten sind Männer ohne Ausbildung hinsichtlich ihrer Chancen zur Familiengründung mehr betroffen als Männer mit akademischer Qualifikation.

Andererseits sind Männer in geringerem Maße auf eine bestimmte Lebensphase fixiert, deshalb können sie sich den historischen Umständen leichter anpaßen. Dies wird an der historischen Variabilität der Wartezeit bis zur Familiengründung von Akademikern ersichtlich. Eine interessante, historisch einmalige Konstellation ergab sich bei der Kohorte 1939-41, bei der die familialen Lebensverläufe von Männern und Frauen von der gewöhnlichen geschlechtsspezifischen Differenzierung abwichen. Bei dieser Kohorte ist der Fall eingetreten, daß durch die

akademische Bildung Frauen sehr viel länger gewartet haben als
Männer, mit der Folge, daß sie in ungefähr dem gleichen
Lebensalter wie Männer geheiratet haben. Wie sich dieses Faktum
auf die Familienstruktur ausgewirkt hat, wird noch zu untersuchen
sein.

Literaturverzeichnis

ANESHENSEL, C.S./ROSEN, B.C. (1980):
   Domestic roles and sex differences in occupational
   expectations
   In: Journal of Marriage and the Family,42, 121-131

BECKER, G. (1973):
   A Theory of Marriage. In: Schultz,Th.W. (ed.): Economics
   of the Family. Chicago/London. The University of Chicago
   Press

BLOSSFELD, H.-P. (1985): Berufseinstieg und
   Segregationsprozess-
   Eine Kohortenanalyse über die Herausbildung von
   geschlechtsspezifischen Berufschancen. SFB 3 Arbeitspapier
   Nr. 171.

BLOSSFELD, H.-P. (1986): Berufseinstieg und Berufsverlauf -
   Eine Kohortenanalyse über die Bedeutung des ersten Berufs
   in der Erwerbsbiografie. SFB 3 Arbeitspapier 161.

BRÜCKNER, E. et al. (1984):
   Methodenbericht "Lebensverläufe". Mannheim. ZUMA.

CARLSON, G./KARLSON, K. (1970):
   Age, Cohorts, and the Generation of Generations.   In:
   American Sociological Review,51,323-341.

COLEMAN, J.S. (1984): Stochastic Models for Market
   Structures. In: Diekmann, A./Mitter, P. (eds): Stochastic
   modelling of social processes. New York: Academic Press

DIEKMANN, A./MITTER, P. (1984): Methoden zur Analyse von
   Zeitverläufen. Stuttgart; Täubner.

FEATHERMAN, D.L., HAUSER, R.M. (1976): Sexual inequalities
   and socioeconomic achievement in the US, 1962-1973. In:

American Sociological Review, 41, 462-483

FEATHERMAN, D.L./LERNER, R. M. (1985): Ontogenesis and
    Sociogenesis: Problematics for Theory andResearch about
    Development and Socializationn across the Lifespan.
    In: American Sociological Review, 50, 659-676.

GOLDSCHEIDER, F.E., WAITE, L.J. (1985): Sex differences in
    the transition to marriage: Evidence about change.
    Manuskript

GOODE, W.J. (1963): World Revolution and Family Patterns.
    Glencoe: Free Press

HAGESTAD, G.O./NEUGARTEN, B.L. (1985): Age and the Life
    Course. In: Binstock, R.H./Shanas, E. (eds.): Handbook of
    aging and the social sciences, Chap. 2, pp. 35-61.

JÜRGENS,H.W. (Hrsg.) (1973): Partnerwahl und Ehe - Theorie
    und Praxis. Hamburg.

KALBFLEISCH, J.D./PRENTICE, R.L. (1980): The Statistical
Analysis of Failure Time Data. New York: Wiley and Sons.

KANTER, R.M. (1977): Men and Women of the Corporation.
    New York: Basic Books

KÖNIG, R. (1974): Materialien zur Soziologie der Familie.
    Köln.

LAPPE, L. (1986): Auflösung geschlechtsspezifischer
    Arbeitsmarktbarrieren? Zur Integration quantitativer und
    qualitativer Arbeitsmarktforschung. Hektographiertes
    Manuskript, Max-Planck-Institut für Bildungsforschung,
    Berlin.

LÜSCHER, K. (1985): Moderne familiale Lebensformen als
    Herausforderung der Soziologie. In: Lutz,B.(Hrsg.):

Soziologie und gesellschaftliche Entwicklung.
Verhandlungen des 22. Deutschen Soziologentages. New
York/Frankfurt: Campus.

MARINI, M.M. (1980): Sex Differences in the Process of
   Occupational Attainment: A Closer Look.
   In: Soc. Sci. Res., 9: 307-361

MARINI, M.M. (1978): Sex Differences in the Determination of
   Adolescent Aspirations: A Review of Research.
   In: Sex Roles, 4: 723-753

MCCLENDON, M.J. (1976): The Occupational Status Attainment
   Process of Males and Females.
   In: Am. Sociol. Rev., 41: 52-64

MEYER, S., SCHULZE, E. (1983): Nicht-eheliche Lebens-
   gemeinschaften - Alternative zur Ehe? In: KZfSS, 4:735-754

MODELL, J. (1980): Normative Aspects of American Marriage
   Timing Since World War II.
   In: Journal of Family History,5,210-234

MODELL, J., FÜRSTENBERG, F.F., STRONG, M. (1978): The Timing
   of Marriage in the Transition to Adulthood: Continuity and
Change 1860-1975. In: American Journal of Sociology, 84,

NAVE-HERZ, R. (1967): Soziologische Aspekte der Frühehe.
   In: KZfSS,19,484-510.

NEIDHARDT, F. (1975): Die Familie in Deutschland.
   Gesellschaftliche Stellung, Struktur und Funktion.
   Opladen. Leske + Budrich.

NEUGARTEN, B.L./DATAN, N. (1973): Sociological Perspecitves

on the Life Cycle. In: Baltes, P.B./Schaie, K.W. (eds.):
Life-Span Developmental Psychology: Personality and
Socialization. Academic Press.

PAPASTEFANOU, G. (1986): Veränderungen des Heiratsalters in
der Bundesrepublik Deutschland seit dem Zweiten Weltkrieg.
Unveröffentliches Manuskript. Max-Planck-Institut für
Bildungsforschung.

PETERSON, T. (1986): Estimating Fully-parametric Hazard Rate
Models with Time-Depenent Covariates by the Method of
Maximum Likelihood. forthcoming: Sociological Methods and
Research.

PETERSON, T. (1984): Presenting Results form Continous Time
Hazard Rate Modells. Manuskript.

RINDFUSS, R.R./BUMPASS, L./JOHN,C.ST. (1980): Education and
Fertility: Implications for the Roles Women Occupy.
In: American Sociological Review,45,431-447.

ROSENBAUM, H. (1978): Familie als Gegenstruktur zur
Gesellschaft. Stuttgart: Enke

ROSENFELD, R. A. (1983): Sex Segregation and Sectors. An
Analysis of Gender Differences in Return from Employer
Changes. In: American Sociological Review,48,637-655.

RÜCKERT, G.-R./LENGSFELD, W./HENKE, W. (1979): Partnerwahl.
Schriftenreihe für Bevölkerungsforschung, Bd. 5

SEWELL, W.H./HAUSER, R.M./WOLF, W.C. (1980): Sex, Schooling,
and Occupational Status. In: Am. J. Sociol., 8:551-583

SOERENSEN, A.B./SOERENSEN, A. (1984): An Event History
Analysis of the Process of Entry into First Marriage.
Wisconsin, Mad. Manuskript.

STATISTISCHES BUNDESAMT (1964): Jugend im wirtschaftlichen

und sozialen Leben der BRD. In: Statistik der BRD.
Stuttgart/Mainz

TREIMAN, D.J./TERELL (1975): Sex and the Process of Status
Attainment. A Comparison of Working Women and Men.
In: American Sociological Review, 40:174-200

WAITE, L.J./SPITZE, G. (1981): Young Women's Transition to
Marriage. In: Demography, 18,

WIEDENBECK, M. (1982): Zum Problem repräsentativer
Querschnitte von kleinen Teilgruppen der Bevölkerung am
Beispiel des Projekts "Lebensverläufe". In: ZUMA -
Nachrichten,10, 21-34.

Übersicht 1

Codierung der Prädiktoren

Prädiktor	Codierung
Kohorte 1939/41	1: Geburtsjahr zwischen 1939 und 1941 0: Anderes Geburtsjahr
Kohorte 1949/51	1: Geburtsjahr zwischen 1949 und 1951 0: Anderes Geburtsjahr
Geschlecht	1: Männlich 0: Weiblich
Kleinstadt	1: Wohnort vor der Eheschließung ist als Kleinstadt bezeichnet 0: andere Bezeichnung des Wohnortes
Mittelstadt	1: Wohnort vor der Eheschließung ist als Mittelstadt bezeichnet 0: andere Bezeichnung des Wohnortes
Großstadt	1: Wohnort vor der Eheschließung ist als Großstadt bezeichnet
Konfession	1: Protestantisch 0: katholisch
Realschule	1: Person hat Realschulabschluß 0: Person hat einen anderen Schulabschluß
Abitur	1: Person hat Abitur 0: Person hat einen anderen Schulabschluß
Hauptschule ohne Abschluß	1: Person hat keinen Hauptschulabschluß 0: Person hat einen anderen Schulabschluß
Angelernt	1: Person hat eine zweijährige Anlernzeit abgeschlossen oder den Abschluß für den einfachen Beamtendienst oder eine Teilabschnittsprüfung oder keinen Abschluß, da die Ausbildung kein formaler Ausbildungsgang war (z.B. Praktikum) 0: Person hat einen anderen beruflichen Abschluß
Fachschule	1: Person hat den Abschluß einer Fach bzw. Berufsfachschule, oder den Abschluß im mittleren oder gehobenen Beamtendienst oder den Abschluß einer beruflichen Weiterbildung als Meister 0: Person hat einen anderen beruflichen Ausbildungsabschluß
Universität	1: Person hat Universitätsabschluß oder einen Abschluß im höheren Dienst 0: Person hat einen anderen beruflichen Abschluß
Kein beruflicher Abschluß	1: Personen hat die Ausbildung abgebrochen oder ist zum Zeitpunkt der Befragung noch in Ausbildung 0: Person hat einen Abschluß bzw. hat nie eine Ausbildung begonnen
Keine berufliche Ausbildung	1: Person hat bis zum Befragungszeitpunkt keine Ausbildung begonnen 0: Person hat eine Ausbildung begonnen
Mittlerer Status	1: Statusscore (1) des ersten Berufes liegt zwischen 77 und 168 0: anderer Statusscore
Hoher Berufsstatus	1: Statusscore (1) des ersten Berufes ist größer als 168 0: anderer Statusscore
Kein beruflicher Status	1: Person war vor der Heirat nicht berufstätig 0: Person war vor der Heirat berufstätig

Anmerkung:

Statusscore ist nach MAYER, K.U. (1977) gebildet worden

Anhang: Tabellen

Tabelle A-1:

Mittelwerte der Prädiktoren im Modell

Prädiktor	Mittelwert
Kohorte B	.33
Kohorte C	.34
Geschlecht	.51
Kleinstadt	.20
Mittelstadt	.15
Großstadt	.31
Konfession	.52
Realschule	.15
Abitur	.09
Hauptschule ohne Abschluß	.08
Angelernt	.03
Fachschule	.05
Universitaet	.04
kein berufl. Abschluß	.22
keine berufl. Ausbildung	.15
mittlerer Berufsstatus	.54
hoher Berufsstatus	.06
kein berufl. Status	.07
Zahl der Fälle	2039

Tabelle A-2

Beta-Gewichte und Standardfehler der Prädiktoren der Heiratswahrscheinlichkeit ab Heiratsmündigkeit, Vollmodell für Männer und Frauen

Prädiktor	Frauen	Männer
Konstante	-13.527 / 0.398	-10.117 / 0.524
Alter	2.055 / 0.074	2.843 / 0.105
Geburtsjahrgänge 1939-1941	0.665 / 0.148	0.125 / 0.140
Geburtsjahrgänge 1949-1951	1.092 / 0.148	0.165 / 0.146
Kleinstadt	0.247 / 0.160	0.024 / 0.166
Mittelstadt	-0.021 / 0.190	0.084 / 0.180
Großstadt	-0.314 / 0.147	-0.314 / 0.142
Konfession	0.017 / 0.115	0.013 / 0.113
Realschule	-0.533 / 0.175	-0.777 / 0.173
Abitur	-1.112 / 0.296	-1.741 / 0.288
Hauptschule ohne Abschluß	-0.054 / 0.219	-0.064 / 0.204
Angelernt	-0.484 / 0.437	-0.180 / 0.365
Fachschule	-0.390 / 0.223	0.305 / 0.501
Universität	-1.135 / 0.441	-0.206 / 0.361
keine berufl. Ausbildung	-0.026 / 0.182	-1.324 / 0.204
kein berufl. Abschluß	-0.149 / 0.183	-0.076 / 0.176
mittlerer Berufsstatus	-0.264 / 0.170	0.205 / 0.146
hoher Berufsstatus	-0.577 / 0.343	-0.012 / 0.304
kein berufl. Status	1.059 / 0.224	1.238 / 0.329
Log Likelihood	-4954.762	-5011.300
Anteil der zens. Fälle	6.0%	11.3%
Anzahl der Fälle	1004	1035

Tabelle A-3

Beta-Gewicht und Standardabweichung von Prädiktoren der Heiratswahrscheinlichkeit ab Heiratsmündigkeit, Basismodell für Männer und Frauen

Prädiktor	Männer	Frauen
Konstante	-13.647 / .383	-12.593 / .341
Alter	1.920 / .078	1.768 / .070
Geburtsjahrgänge 1939-1941	0.266 / .145	0.450 / .147
Geburtsjahrgänge 1949-51	0.107 / .138	0.721 / .138
Anzahl der Fälle	1004	1035
Zensierung in Prozent	11.3	6.0
Log Likelihood	-4617.1	-4680.5

Tabelle A-4

Beta-Gewicht und Standardabweichung von Prädiktoren der Heirats wahrscheinlichkeit Basismodell der Geburtsjahrgänge 1929-31, 1939-41 und 1949-51

Prädiktor	Geburtsjahrgänge		
	1929-39	1939-41	1949-51
Konstante	-13.750 / .421	-12.950 / .382	-11.042 / .456
Alter	2.024 / .088	1.953 / .083	1.573 / .101
Geschlecht	-0.381 / .141	-0.581 / .145	-0.911 / .137
Anzahl der Fälle	653	658	675
Zensierung in Prozent	10.3	11.2	22.3
Log Likelihood	-3236.1	-3168.3	-2868.0

Tabelle A-5

Beta-gewichte und Standardabweichung von Praedik-
toren der Heiratswahrscheinlichkeit ab
Heiratsmundigkeit
Vollmodell fur die Kohorten 1929-31, 1939-41
und 1949-51

Pradiktor	Geburtsjahrgaenge		
	1929-31	1939-41	1949-51
Konstante	-13.922 .456	-13.459 .443	-12.002 .504
Alter	2.120 .093	2.117 .090	1.903 .112
Geschlecht	-0.500 .165	-0.602 .166	-0.871 .160
Kleinstadt	0.057 .198	0.195 .207	0.375 .198
Mittelstadt	0.031 .243	0.099 .229	-0.652 .227
Großstadt	-0.143 .179	-0.145 .188	-0.625 .182
Konfession	-0.002 .145	0.067 .151	-0.079 .140
Realschule	-0.548 .260	-0.678 .231	-0.704 .192
Abitur	-0.955 .506	-1.568 .355	-1.646 .334
Hauptschule ohne Abschluß	-0.083 .231	-0.254 .280	-0.328 .307
Angelernt	-0.332 .659	-0.545 .534	-0.406 .449
Fachschule	-0.274 .430	-0.600 .343	0.300 .320
Universitaet	-0.674 .754	-0.700 .514	-0.608 .394
kein berufl. Abschluß	-0.081 .229	-0.160 .224	-0.302 .216
keine berufl. Ausbildung	-0.303 .207	-0.053 .206	-0.038 .246
mittlerer Berufsstatus	0.291 .237	-0.077 .196	0.105 .190
kein berufl. Status	0.843 .343	1.678 .364	0.972 .273
Zahl der Falle	653	658	675
Zensierung in Proz.	10.3	11.2	22.4
Log Likelihood	-3217.4	-3134.8	-2817.9

Tabelle A-6

Beta-Gewicht und Standardabweichung von Prädiktoren der
Heiratswahrscheinlichkeit ab Heiratsmündigkeit
Basismodell für Frauen der Kohorten 1929-31, 1939-41 und
1949-51

Prädiktor	Geburtsjahrgänge		
	1929-39	1939-41	1949-51
Konstante	-13.398 .625	-12.299 .481	-11.087 .579
Alter	1.946 .132	1.804 .106	1.594 .129
Anzahl der Fälle	327	316	340
Zensierung in Prozent	8.9	11.1	15.3
Log Likelihood	-1633.8	-1507.8	-1537.1

Tabelle A-7

Beta-Gewichte und Standardabweichung von Prädiktoren
der Heiratswahrscheinlichkeit ab Heiratsmündigkeit
Vollmodell für Frauen der Kohorten 1929-31, 1939-41
und 1949-51

Pradiktor	Geburtsjahrgaenge		
	1929-31	1939-41	1949-51
Konstante	-13.860 .689	-13.362 .608	-12.095 .620
Alter	2.072 .137	2.099 .122	1.907 .150
Kleinstadt	0.434 .280	0.007 .306	0.347 .275
Mittelstadt	-0.131 .396	-0.407 .383	0.054 .320
Großstadt	-0.181 .255	-0.510 .285	-0.519 .269
Konfession	-0.029 .204	0.185 .230	-0.112 .201
Realschule	-0.436 .346	-0.502 .352	-0.777 .259
Abitur	-0.745 .763	-1.434 .477	-1.341 .529
Hauptschule ohne Abschluß	0.094 .340	-0.242 .393	-0.277 .475
Angelernt	-0.689 1.041	-0.389 .895	-0.678 .609
Fachschule	-0.421 .465	-0.773 .392	0.273 .362
Universitaet	-0.498 1.470	-1.213 .755	-1.133 .625
kein berufl. Abschluß	-0.095 .383	0.050 .328	-0.403 .273
keine berufl. Ausbildung	-0.059 .298	0.274 .281	0.381 .293
mittlerer Berufsstatus	-0.270 .539	-0.136 .421	0.250 .368
kein berufl. Status	0.862 .382	1.710 .407	1.121 .346
Zahl der Falle	327	316	340
Zensierung in Proz.	8.9	11.1	15.3
Log Likelihood	-1622.4	-1478.6	-1500.7

Tabelle A-8

Beta-Gewicht und Standardabweichung von Prädiktoren
der Heiratswahrscheinlichkeit ab Heiratsmündigkeit
Basismodell für Männer der Geburtsjahrgänge 1929-31,
1939-41 und 1949-51

Prädiktor	Geburtsjahrgänge		
	1929-39	1939-41	1949-51
Konstante	-14.543	-14.204	-11.886
	.567	.648	.767
Alter	2.113	2.101	1.556
	.118	.135	.162
Anzahl der Fälle	326	342	335
Zensierung in Prozent	11.7	11.4	29.6
Log Likelihood	-1602.0	-1659.5	-1350.9

Tabelle A-9

Beta-Gewichte und Standardabweichung von Praediktoren
der Heiratswahrscheinlichkeit ab Heiratsmündigkeit.
Vollmodelle für Männer der Kohorten 1929-31,1939-41 und
1949-51

Prädiktor	Geburtsjahrgaenge		
	1929-31	1939-41	1949-51
Konstante	-14.951	-14.750	-13.143
	.629	.715	.855
Alter	2.257	2.279	1.935
	.134	.149	.184
Kleinstadt	-0.273	0.270	0.456
	.255	.289	.288
Mittelstadt	0.262	0.601	-0.206
	.271	.303	.320
Großstadt	-0.080	0.123	-0.791
	.227	.256	.253
Konfession	0.088	0.036	-0.016
	.212	.212	.200
Realschule	-0.645	-0.943	-0.705
	.425	.324	.287
Abitur	-1.157	-2.179	-1.937
	.691	.589	.487
Hauptschule ohne Abschluß	-0.178	0.087	-0.209
	.321	.395	.464
Angelernt	-0.002	-1.039	-0.356
	1.018	.675	.663
Fachschule	0.035	0.345	0.398
	1.710	.857	.741
Universitaet	-0.957	0.468	-0.159
	.893	.804	.515
kein berufl. Abschluß	-0.112	-0.393	-0.124
	.290	.320	.354
keine berufl. Ausbildung	-1.044	-1.482	-2.500
	.312	.351	.600
mittlerer Berufsstatus	0.314	-0.227	-0.072
	.278	.231	.234
kein berufl. Status	0.499	1.628	0.666
	.959	.876	.471
Zahl der Fälle	326	342	335
Zensierung in Proz.	11.7	11.4	29.6
Log Likelihood	-1585.3	-1640.7	-1306.6

Tabelle A-10

Mittelwerte der Prädiktoren der Heiratsrate
von Frauen der Jahrgänge 1929-31, 1939-41 und
1949-51

Prädiktor	Geburtsjahrgaenge		
	1929-31	1939-41	1949-51
Kleinstadt	19.2	21.2	23.8
Mittelstadt	12.2	13.6	15.9
Großstadt	24.8	29.7	32.9
Konfession	50.2	54.7	51.2
Realschule	13.1	15.2	18.5
Abitur	5.2	7.0	10.3
Hauptschule ohne Abschluß	9.2	6.6	5.9
Angelernt	1.8	2.2	2.4
Fachschule	6.4	9.2	9.7
Universitaet	1.5	4.1	6.5
keine berufl. Ausbildung	54.4	36.1	16.8
kein berufl. Abschluss	13.8	13.6	16.5
mittlerer Berufsstatus	5.5	7.6	9.1
kein berufl. Status	10.4	7.6	9.7
Zahl der Fälle	327	316	340

Tabelle A-11

Mittelwerte der Prädiktoren der Heiratsrate von
Männer der Kohorten 1929-31, 1939-41 und 1949-51

Prädiktor	Geburtsjahrgaenge		
	1929-31	1939-41	1949-51
Kleinstadt	26.1	19.9	18.2
Mittelstadt	20.9	13.4	16.1
Großstadt	35.3	28.9	33.4
Konfession	45.4	53.8	54.9
Realschule	10.4	13.7	18.5
Abitur	7.1	5.6	16.4
Hauptschule ohne Abschluß	10.7	7.0	6.7
Angelernt	1.5	1.5	4.5
Fachschule	2.1	2.3	3.0
Universitaet	2.1	2.6	9.6
keine berufl. Ausbildung	15.3	8.5	3.0
kein berufl. Abschluss	20.2	10.5	12.8
mittlerer Berufsstatus	29.1	42.7	37.3
kein berufl. Status	2.8	2.0	8.4
Zahl der Fälle	326	342	335

Tabelle 1

Prädiktoren der Heiratswahrscheinlichkeit ab Heiratsmündigkeit
Schrittweise Erweiterung des Modells
Beta Gewichte und Standardfehler der loglogistischen
Verteilungsschätzung

Prädiktor	\ Modell 1	2	3	4	5	6	7
Log Likelihood	10285	10280	10120	10120	10054	10048	10010
Konstante	13.258	13.879	13.947	13.948	14.404	14.319	14.605
	.239	.260	.266	.268	.272	.282	.295
Dauer	1.889	2.115	2.159	2.159	2.281	2.288	2.338
	.050	.054	.055	.055	.056	.057	.058
Kohorte 1939 41		.331	.358	.358	.356	.354	.382
		.099	.099	.099	.100	.101	.101
Kohorte 1949 51		.462	.496	.496	.631	.623	.623
		.094	.095	.096	.098	.101	.102
Männer		1.256	1.281	1.281	1.300	1.370	1.332
		.083	.083	.083	.084	.091	.092
Kleinstadt			.119	.119	.203	.187	.200
			.114	.114	.114	.115	.115
Mittelstadt			.073	.073	.116	.107	.066
			.126	.126	.128	.129	.130
Großstadt			.512	.512	.235	.260	.259
			.096	.097	.099	.102	.102
Protestanten				.005	.003	.011	.018
				.079	.080	.080	.080
Realschule					.681	.655	.676
					.117	.119	.121
Abitur					1.501	1.199	1.443
					.154	.197	.200
Hauptschule ohne Abschluß					.186	.168	.121
					.147	.149	.149
Angelernt						.371	.340
						.282	.291
Fachschule						.302	.232
						.168	.189
Universität						.695	.690
						.262	.279
Kein Abschluß						.182	.230
						.110	.129
Keine Ausbildung						.053	.125
						.117	.126
Mittlerer Status							.015
							.111
Hoher Status							.233
							.225
nicht erwerbstätig							1.193
							.179

Tabelle 2
Anteile von Personen, die zwischen dem 30. und
dem 40. Lebensjahr heiraten und ihre mittlere
"Wartezeit" bis zur ersten Eheschließung;
nach sozialstrukturellen Merkmalen differenzierte
Schätzwerte

Soziale Gruppe	Anteile der Heiratenden	Mittlere Wartezeit
Kohorte 1929/31	21.5	17.6
Kohorte 1939/41	16.5	17.4
Kohorte 1949/51	13.6	17.5
Frauen	6.5	17.3
Männer	17.0	17.4
Dorf	16.4	17.7
Großstadt	20.0	17.5
Hauptschüler	15.2	17.4
Realschüler	16.2	17.4
Abiturienten	35.2	18.3
Lehrabschluß	15.5	17.5
Universität	26.0	17.6
keine Ausb.	18.4	17.5
kein Abschl.	17.0	17.6
mittl. Status	17.9	17.7
keine Erwerbst.	6.7	17.4

Tabelle 3

Mittlere Wartezeit bis zur ersten Eheschließung
von Personen, die bis zum 50. Lebensjahr
geheiratet haben, in Jahren ab Heiratsmündigkeit,
nach sozialen Gruppen, Vollmodell

Soziale Gruppe	Mittlere Wartezeit
Kohorte 1929/31	11.5
Kohorte 1939/41	10.4
Differenz	1.1
Kohorte 1949/51	9.8
Differenz	0.6
Frauen	7.3
Männer	10.5
Differenz	3.2
Dorf	10.5
Großstadt	11.2
Differenz	0.7
Hauptschule	10.1
Realschule	12.3
Differenz	2.2
Abiturienten	14.6
Differenz	4.5
Lehre	10.2
Universität	12.3
Differenz	2.1
keine Ausbildung	10.9
Differenz	0.7
kein Abschluß	10.6
Differenz	0.4
niedriger Status	10.8
nicht erwerbstätig	7.8
Differenz	2.0

Tabelle 4

Mittlere Wartezeit bis zur ersten Eheschließung
von Personen, die bis zum 50. Lebensjahr
geheiratet haben, in Jahren ab Heiratsmündigkeit,
nach sozialen Gruppen, reduziertes Modell

Soziale Gruppe	Mittlere Wartezeit
Kohorte 1929/31	9.9
Kohorte 1939/41	9.0
Differenz	0.9
Kohorte 1949/51	8.7
Differenz	0.3
Frauen	7.6
Männer	10.9
Differenz	3.3

Tabelle 5

Mittelwerte der Prädiktoren bei Männern und Frauen

Prädiktor	Mittelwerte	
	Frauen	Männer
Kohorte 1939/41	.318	.345
Kohorte 1949/51	.344	.333
Kleinstadt	.210	.193
Mittelstadt	.144	.148
Großstadt	.298	.324
Protestanten	.519	.518
Realschule	.152	.143
Abitur	.075	.100
Hauptschule ohne Abschl.	.076	.080
Angelernt	.022	.028
Fachschule	.083	.025
Universität	.039	.046
keine Ausbildung	.353	.087
kein berufl. Abschl.	.143	.147
Mittlerer Status	.505	.584
Hoher Status	.055	.062
kein Berufstatus	.094	.044
Zahl der Fälle	1004	1035

Tabelle 6

Mittlere Wartezeit bis zur ersten Heirat
von Frauen und Männern, die bis zum 50.
Lebensjahr geheiratet haben
in Jahren ab Heiratsmündigkeit, nach sozialen
Gruppen

soziale Gruppe	Männer	Frauen
Kohorte 1929/31	11.0	9.0
Kohorte 1939/41	10.7	7.4
Differenz	0.3	1.6
Kohorte 1949/51	10.6	6.4
Differenz	0.1	1.0
Dorf	10.5	7.6
Großstadt	11.4	8.4
Differenz	0.9	0.8
Hauptschüler	10.0	7.2
Realschüler	12.1	8.4
Differenz	2.1	1.2
Abitur	17.6	10.0
Differenz	7.6	2.8
Lehre	10.4	7.3
Uni	10.9	10.3
Differenz	0.5	3.0
keine Ausb.	14.2	7.3
Differenz	3.8	0.0
kein Abschl.	10.6	7.7
Differenz	0.2	0.4
Niedr.Status(1)	11.2	7.4
Hoher Status(1)	11.2	8.8
Differenz	0.0	1.4
Kein Status(1)	8.3	5.3
Differenz	2.9	2.1

Tabelle 7

Mittlere Wartezeit bis zur ersten Heirat
von Frauen und Männern, die bis zum 50.
Lebensjahr geheiratet haben, reduziertes Modell,
in Jahren ab Heiratsmündigkeit, nach sozialen
Gruppen

soziale Gruppe	Männer	Frauen
Kohorte 1929/31	10.1	9.1
Kohorte 1939/41	9.3	7.9
Differenz	0.8	1.2
Kohorte 1949/51	9.8	7.2
Differenz	0.5	0.7
Durchschnitt	9.7	8.0

Tabelle 8

Mittlere Wartezeit bis zur ersten Heirat von
Personen, die bis zum 50. Lebensjahr geheiratet
haben.
Jahrgänge 1929/31, 1939/41 und 1949/51,
reduziertes Modell, in Jahren ab Heiratsmündigkeit

soziale Gruppe	Kohorte 1929/31	Kohorte 1939/41	Kohorte 1949/51
Frauen	8.9	7.8	7.3
Männer	10.0	9.3	10.0
Differenz	1.1	1.5	2.7
Durchschnitt	9.5	8.5	8.6

Tabelle 9

Mittlere Wartezeit bis zur ersten Heirat von
Personen, die bis zum 50. Lebensjahr geheiratet
haben
Jahrgänge 1929/31, 1939/41 und 1949/51,
in Jahren ab Heiratsmündigkeit, nach sozialen
Gruppen

soziale Gruppe	Kohorte 1929/31	Kohorte 1939/41	Kohorte 1949/51
Frauen	8.8	7.7	7.3
Männer	10.1	9.2	9.6
Differenz	1.3	1.5	2.3
Dorf	9.4	8.5	8.0
Großstadt	9.8	8.8	9.7
Differenz	0.3	0.3	1.7
Hauptschule	9.1	7.9	7.5
Realschule	10.6	9.7	9.3
Differenz	1.5	1.8	1.8
Abitur	11.9	12.4	12.2
Differenz	2.8	4.5	4.7
Lehre	9.0	8.2	8.2
Universität	10.9	10.0	9.8
Differenz	1.9	1.8	1.6
keine Ausb.	9.9	8.3	8.3
Differenz	0.9	0.1	0.1
kein Abschl.	9.3	8.6	9.0
Differenz	0.3	0.4	0.8
Niedr. Stat.	8.9	8.8	8.4
Nichterwerbst.	7.6	5.2	6.4
Differenz	1.3	3.6	2.0
Durchschnitt Vollmodell	9.4	8.4	8.4
Durchschnitt Reduz.Modell	9.5	8.5	8.6

Tabelle 10

Mittlere Wartezeit bis zur ersten Heirat
von Frauen und Männern, die bis zum 50.
Lebensjahr geheiratet haben, Kohorte 1929/31,
in Jahren ab Heiratsmündigkeit, nach sozialen
Gruppen

soziale Gruppe	Männer	Frauen
Dorf	9.9	9.0
Großstadt	10.1	10.1
Differenz	0.2	1.1
Hauptschule	9.5	8.7
Realschule	11.4	9.5
Differenz	1.9	0.8
Abitur	13.0	10.8
Differenz	3.5	2.1
Lehre	9.4	8.7
Universität	12.2	12.2
Differenz	2.8	3.5
keine Ausb.	12.5	8.9
Differenz	3.1	0.2
kein Abschl.	9.7	9.7
Differenz	0.3	1.0
Niedr. Stat.	10.3	9.1
Nichterwerbst.	8.9	7.1
Differenz	-1.4	-2.0
Durchschnitt Vollmodell	10.0	8.9
Durchschnitt Nullmodell	10.0	9.0

Tabelle 11

Mittlere Wartezeit bis zur ersten Heirat
von Frauen und Männern, die bis zum 50.
Lebensjahr geheiratet haben, Kohorte 1939/41,
in Jahren ab Heiratsmündigkeit, nach sozialen
Gruppen

soziale Gruppe	Männer	Frauen
Dorf	9.6	7.2
Großstadt	9.3	8.4
Differenz	0.3	1.2
Hauptschule	8.6	7.2
Realschule	11.1	8.5
Differenz	2.5	1.3
Abitur	15.2	11.0
Differenz	6.6	3.8
Lehre	8.8	7.6
Universität	7.7	10.9
Differenz	-1.1	3.3
keine Ausb.	13.1	7.0
Differenz	4.3	0.6
kein Abschl.	9.8	9.8
Differenz	1.0	1.8
Niedr. Stat.	9.0	8.0
Nichterwerbst.	5.6	4.7
Differenz	3.4	3.3
Durchschnitt Vollmodell	9.2	7.7
Durchschnitt Nullmodell	9.2	7.9

Tabelle 12

Mittlere Wartezeit bis zur ersten Heirat
von Frauen und Männern der Kohorte 1949/51,
in Jahren nach der Heiratsmündigkeit

soziale Gruppe	Männer	Frauen
Dorf	9.2	7.0
Großstadt	11.6	8.2
Differenz	2.4	1.2
Hauptschule	8.6	6.5
Realschule	10.5	8.3
Differenz	1.9	1.8
Abitur	14.6	9.9
Differenz	6.0	3.4
Lehre	9.5	7.0
Universität	10.1	10.1
Differenz	0.6	3.1
keine Ausb.	17.9	6.2
Differenz	8.4	0.8
kein Abschl.	9.8	9.8
Differenz	0.3	2.8
Niedr. Stat.	9.9	7.4
Nichterwerbst.	8.1	5.2
Differenz	1.8	2.2
Durchschnitt Vollmodell	9.8	7.1
Durchschnitt Nullmodell	10.0	7.3

Tabelle 13

Mittlere Wartezeit bis zur ersten Heirat von Frauen,
die bis zum 50. Lebensjahr geheiratet haben,
Jahrgänge 1929/31, 1939/41 und 1949/51,
in Jahren ab Heiratsmündigkeit, nach sozialen
Gruppen

soziale Gruppe	Kohorte 1929/31	Kohorte 1939/41	Kohorte 1949/51
Dorf	9.0	7.2	7.0
Großstadt	9.5	8.4	8.2
Differenz	0.5	1.2	1.2
Hauptschule	8.7	7.2	6.5
Realschule	9.5	8.5	8.3
Differenz	1.2	1.3	1.8
Abitur	10.8	11.0	9.9
Differenz	2.1	3.8	3.4
Lehre	8.7	7.6	7.0
Universität	10.1	10.9	10.1
Differenz	1.4	3.3	3.1
keine Ausb.	8.9	7.0	6.2
Differenz	0.2	0.6	0.8
kein Abschl.	9.7	9.8	9.8
Differenz	1.0	1.8	2.8
Niedr. Stat.	9.1	8.0	7.4
Nichterwerbst.	7.1	4.7	5.2
Differnz	2.0	3.3	2.2
Durchschnitt Vollmodell	8.9	7.7	7.1
Durchschnitt Nullmodell	9.0	7.9	7.3

Tabelle 14

Mittlere Wartezeit bis zur ersten Heirat von Männern, die bis zum 50. Lebensjahr geheiratet haben, Jahrgänge 1929/31, 1939/41 und 1949/51, in Jahren ab Heiratsmündigkeit, nach sozialen Gruppen

soziale Gruppe	Kohorte 1929/31	Kohorte 1939/41	Kohorte 1949/51
Dorf	9.9	9.6	9.2
Großstadt	10.1	9.3	11.6
Differenz	0.2	0.3	2.4
Hauptschule	9.5	8.6	8.6
Realschule	11.4	11.1	10.5
Differenz	1.9	2.5	1.9
Abitur	13.0	15.2	14.6
Differenz	3.5	6.6	6.0
Lehre	9.4	8.8	9.5
Universität	12.2	7.7	10.1
Differenz	2.8	-1.1	0.6
keine Ausb.	12.5	13.1	17.9
Differenz	3.1	4.3	8.4
kein Abschl.	9.7	9.8	9.8
Differenz	0.3	1.0	0.3
Niedr. Stat.	10.3	9.0	9.9
Nichterwerbst.	8.9	5.6	8.1
Differenz	1.4	3.4	1.8
Durchschnitt Vollmodell	10.0	9.2	9.8
Durchschnitt Nullmodell	10.0	9.2	10.0

Tabelle 15

Mittlere Wartezeit bis zur ersten Heirat Männer und Frauen der Kohorten 1929/31, 1939/41 und 1949/51, bei Angleichung von Kompositions und Alterseffekten der Frauen an die der Männer, in Jahren nach der Heiratsmündigkeit

Angleichung von ...	Kohorte 1929/31		Kohorte 1939/41		Kohorte 1949/51	
	Männer	Frauen	Männer	Frauen	Männer	Frauen
Alterseffekt und Bildungsverteilung	10.0	9.8	9.2	9.3	9.8	9.1
Alterseffekt und Bildungseffekt	10.0	11.2	9.2	9.9	9.8	10.2
Alterseffekt	10.0	9.5	9.2	9.0	9.8	8.7
Bildungseffekt und Bildungsverteilung	10.0	9.4	9.2	9.0	9.8	8.7
Faktische Werte	10.0	8.9	9.2	7.7	9.8	7.1

Tabelle 16

Mittlere Wartezeit bis zur ersten Heirat von Frauen der Kohorten 1939/41 und 1949/51, bei Angleichung an Kompositions und Alterseffekte der Kohorte 1929/31, in Jahren nach der Heiratsmündigkeit

Angleichung von ...	Kohorte 1929/31	Kohorte 1939/41	Kohorte 1949/51
Alterseffekt und Bildungsverteilung	8.9	8.7	8.8
Alterseffekt und Bildungseffekt	8.9	9.1	9.2
Alterseffekt	8.9	9.3	9.7
Bildungseffekt und Bildungsverteilung	8.9	7.4	6.5
Faktische Werte	8.9	7.7	7.1

Tabelle 17

Mittlere Wartezeit bis zur ersten Heirat von Männern der Kohorten 1939/41 und 1949/51, bei Angleichung an Kompositions und Alterseffekte der Kohorte 1929/31, in Jahren nach der Heiratsmündigkeit

Angleichung von	Kohorte 1929/31	Kohorte 1939/41	Kohorte 1949/51
Alterseffekt und Bildungsverteilung	10.0	10.1	11.7
Alterseffekt und Bildungseffekt	10.0	9.6	10.0
Alterseffekt	10.0	10.0	11.4
Bildungseffekt und Bildungsverteilung	10.0	9.2	8.5
Faktische Werte	10.0	9.2	9.8

Schaubild 1
Mittleres Alter von Frauen und Männern bei ersten Eheschließung
1950 bis 1982

Schaubild 1
Mittleres Alter von Frauen und Männern bei ersten Eheschließung 1950 bis 1982

SCHAUBILD 2

HEIRATSRATEN VON FRAUEN UND MAENNERN DES JAHRGANGS 1940

QUELLE: EIGENE ZUSAMMENSTELLUNG AUS DEN STATISTISCHEN JAHRBUECHERN
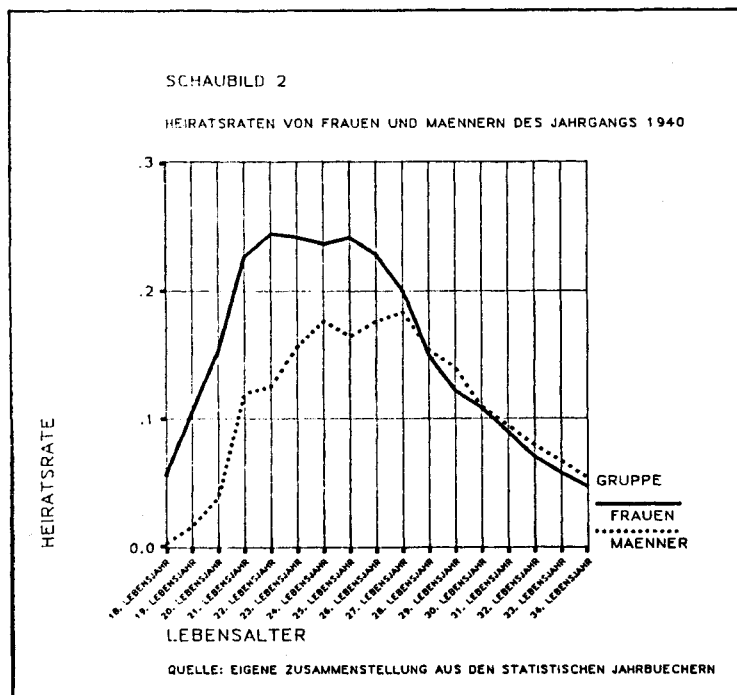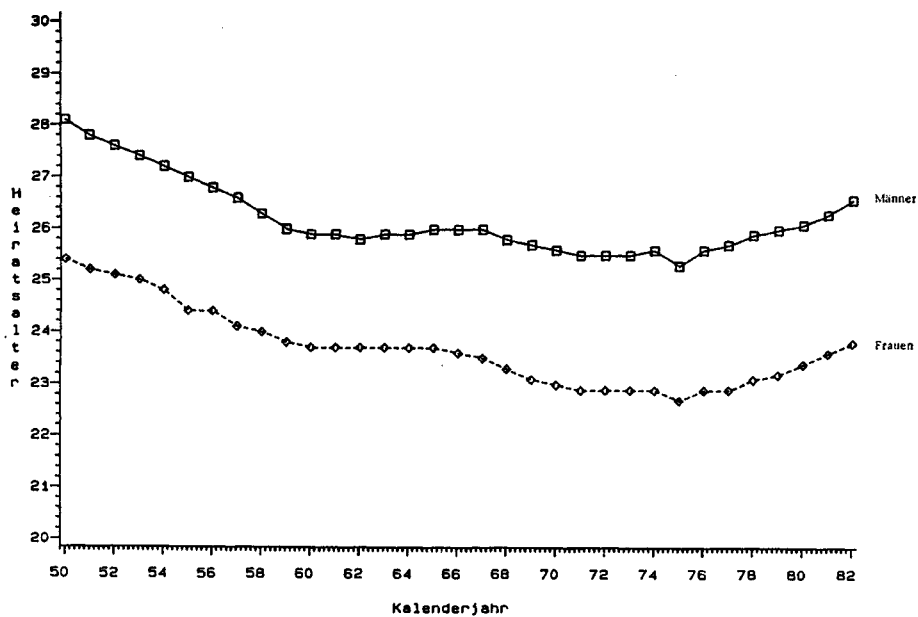
Schaubild 3
Heiratsraten von Frauen der Jahrgänge 1939, 1942, 1951, 1954, 1955, 1958
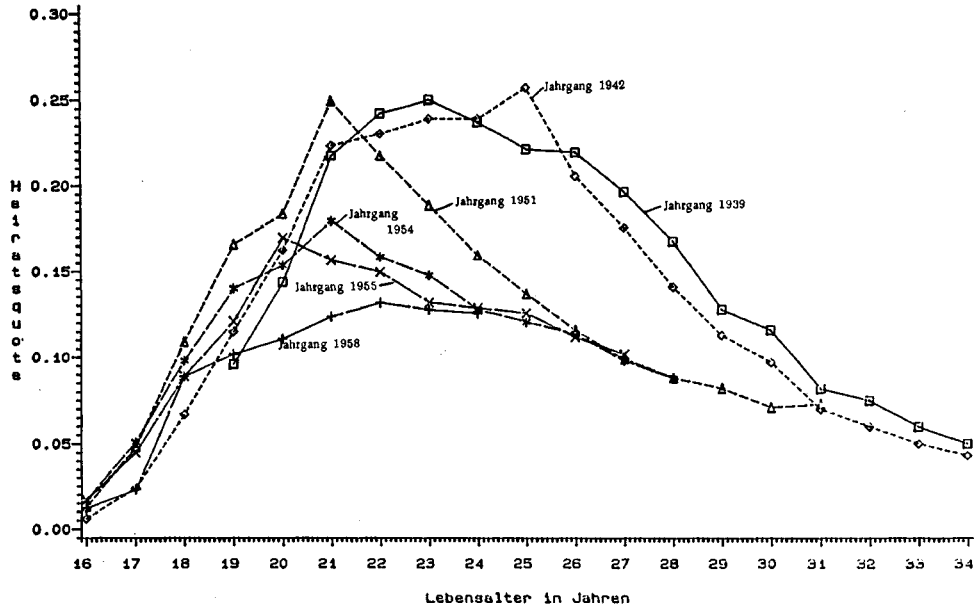
Schaubild 4
Heiratsraten von Frauen im Alter von 16 bis 34 Jahren
Jahrgänge 1929 bis 1959
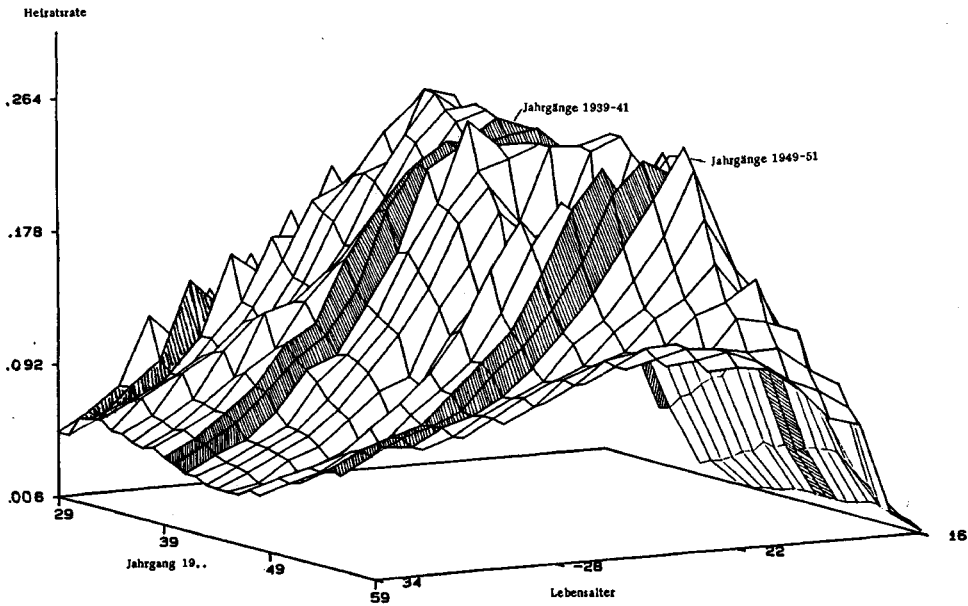
Schaubild 5
Heiratsraten von Männern der Jahrgänge 1939, 1942, 1951, 1954, 1955, 1958
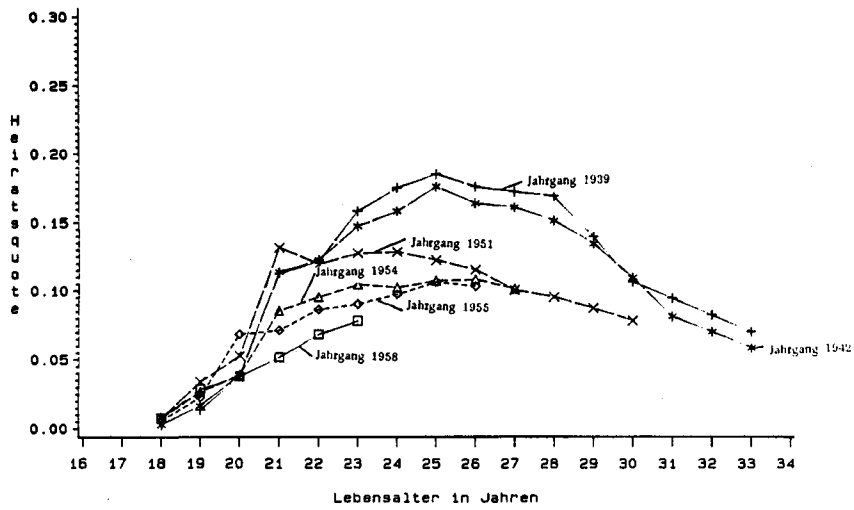
Schaubild 6
Heiratsraten von Männern im Alter von 18 bis 34 Jahren
Jahrgänge 1929 bis 1959
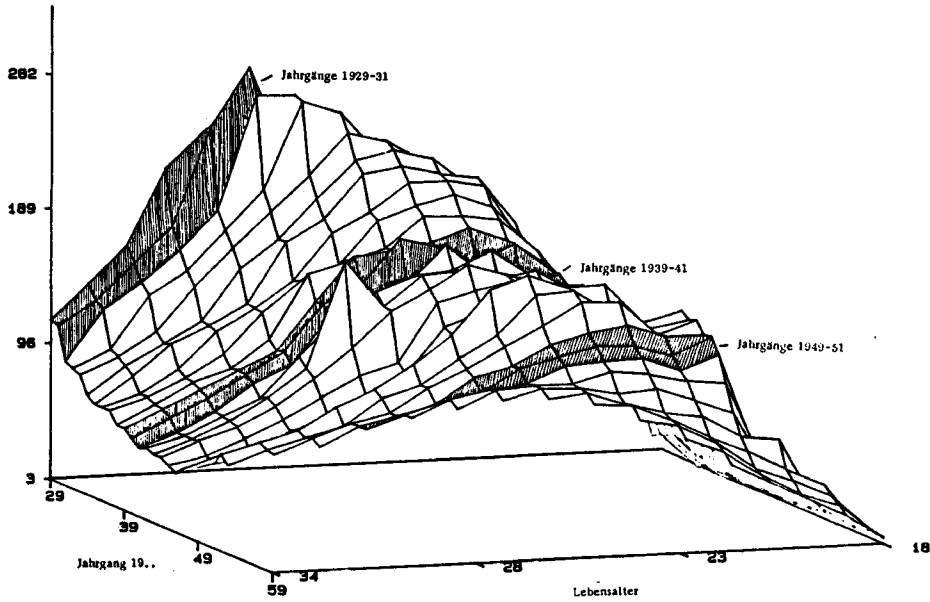
Schaubild 7
Prädizierte Heiratsraten von Frauen und Männern
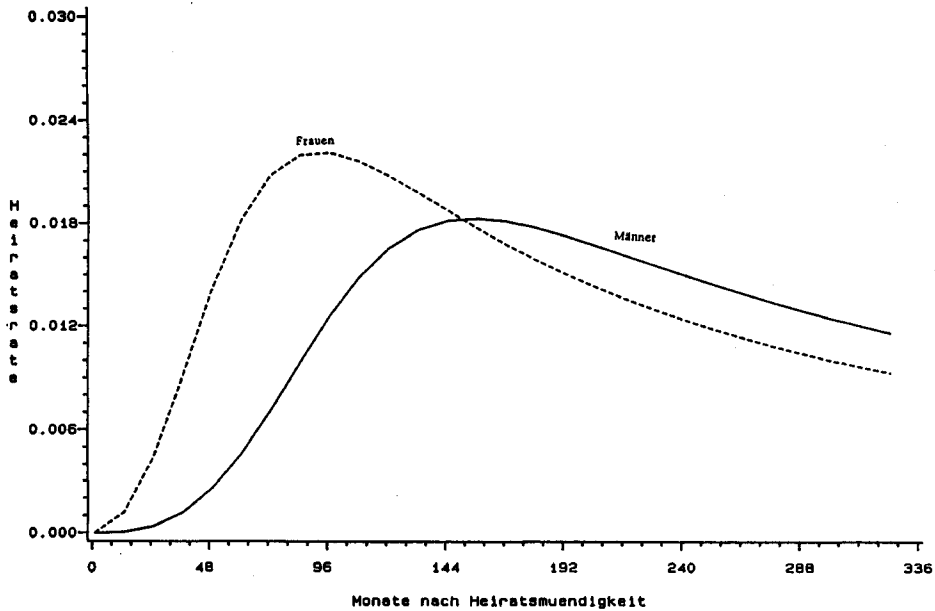
Schaubild 8
Prädizierte Heiratsraten von Frauen und Männern mit Hauptschulabschluß und Lehre
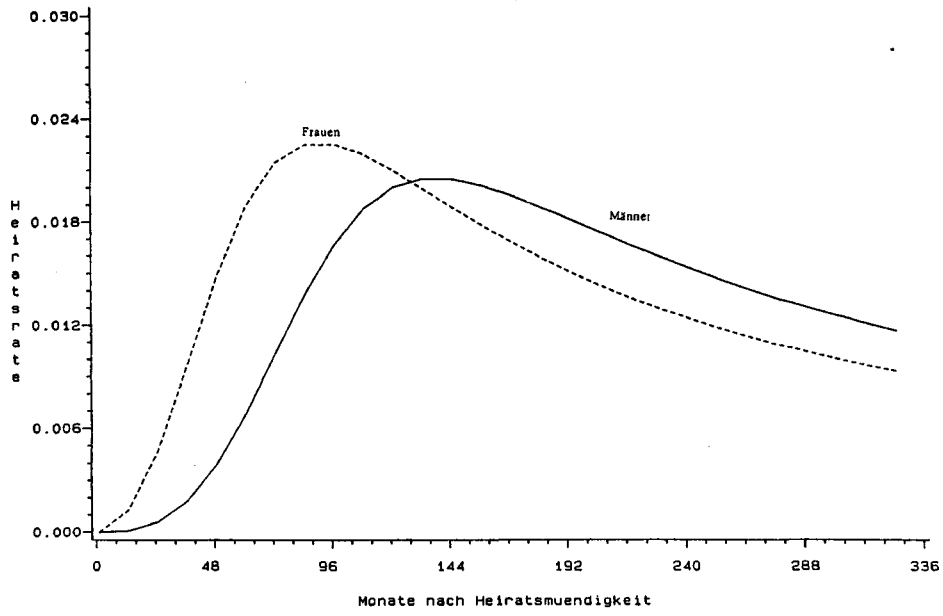
Schaubild 9
Prädizierte Heiratsraten von Frauen und Männern mit akademischer Ausbildung
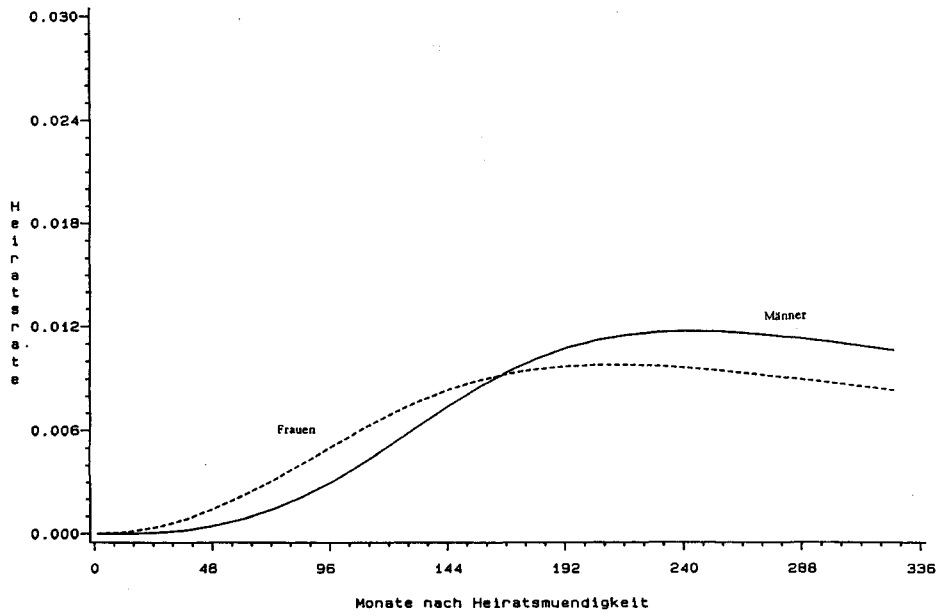
Schaubild 10
Prädizierte Heiratsraten von Frauen und Männern   ohne berufliche Ausbildung

Monate nach Heiratsmuendigkeit
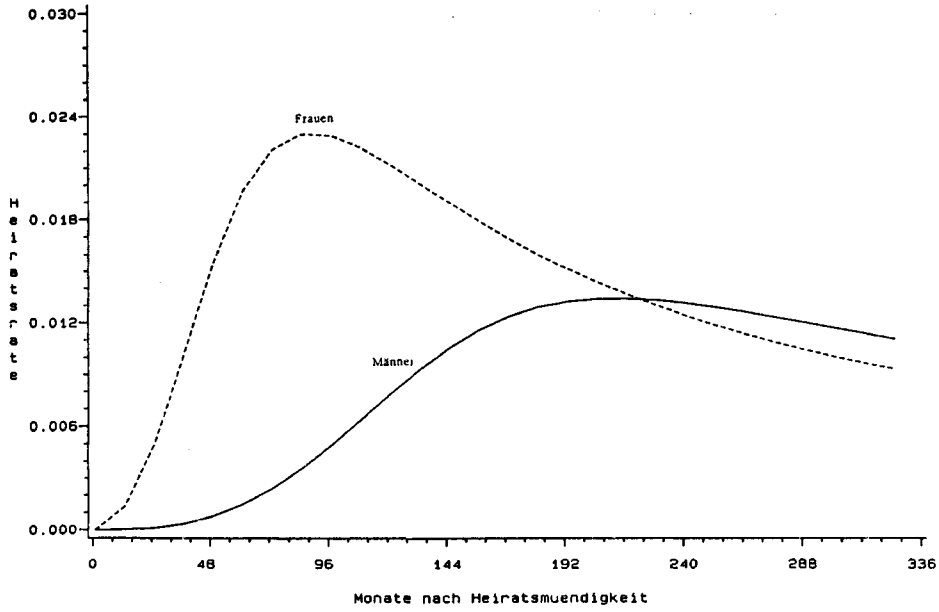
Schaubild 11

Log-logistische Funktion der Hazardrate der ersten Eheschließung mit verschedenen Kovariaten

Modell1: ohne Kovariaten

Modell2: Kohorte und Geschlecht als Kovariaten

Modell3: Kovariaten: Kohorte , Geschlecht, Konfession, Ortstyp,
          allgemeiner Schulabschluß, berufl. Qualifikation.
          Status des ersten Beruf·

.. Monate nach Heiratsmuendigkeit
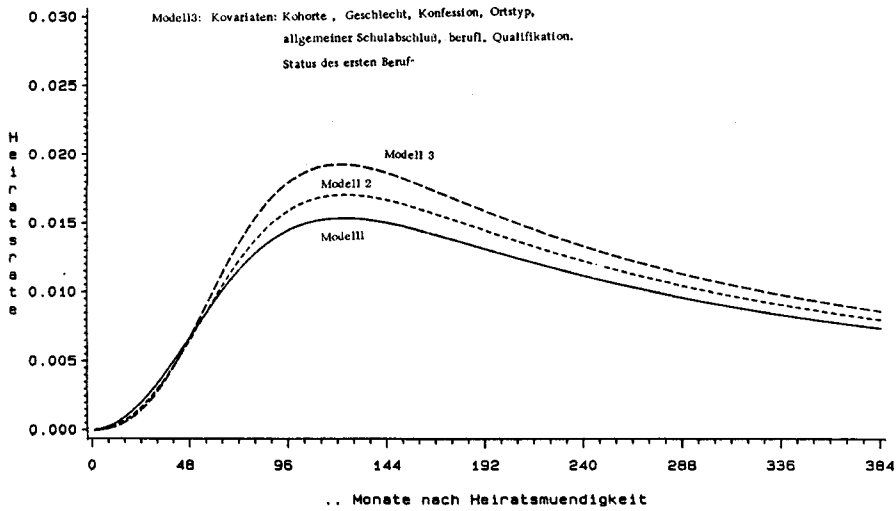
Schaubild 12:
Prädizierte Heiratsraten, Nullmodell und Vollmodell, beide mit Kovariaten

.. Monate nach Heiratsmuendigkeit
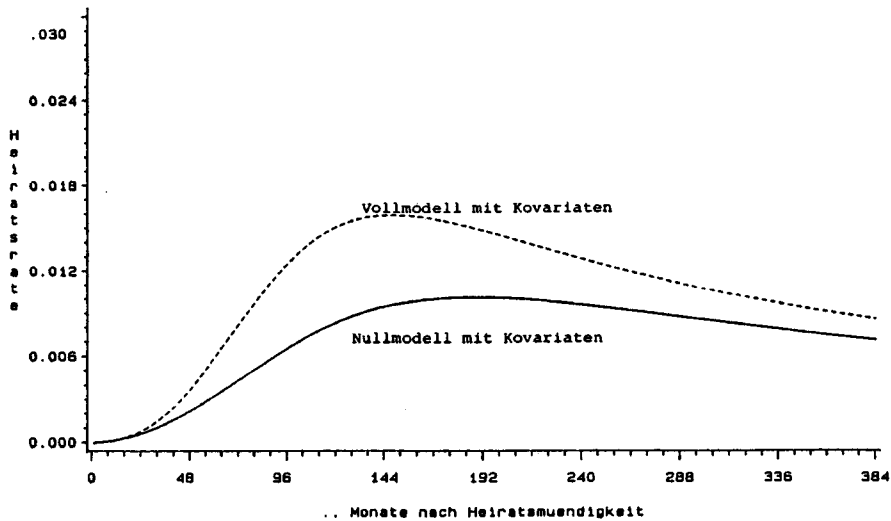
Schaubild 13
Prädizierte Heiratsraten von Männern der Kohorte 1929-31. Basis- und Vollmodell

.. Monate nach Heiratsmuendigkeit
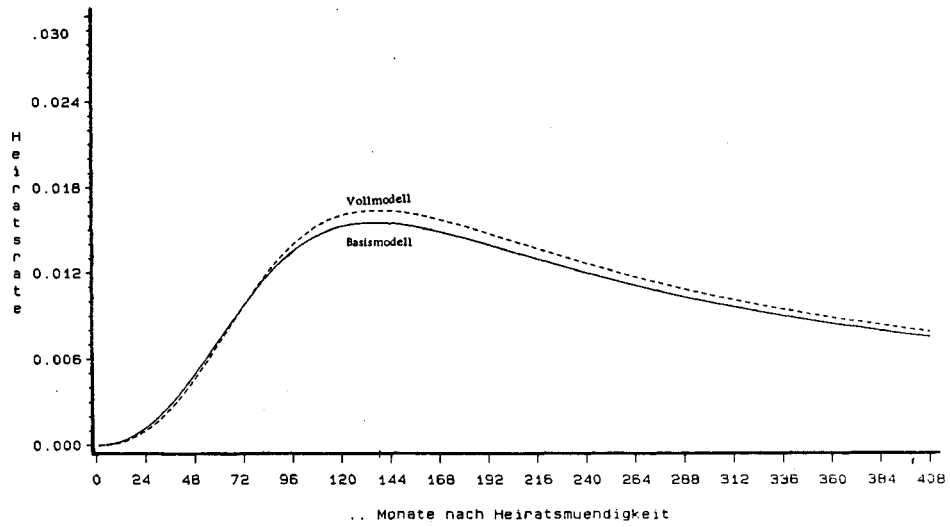
Schaubild 14
Prädizierte Heiratsraten von Männern der Kohorte 1939-41, Basis- und Vollmodell

.. Monate nach Heiratsmuendigkeit
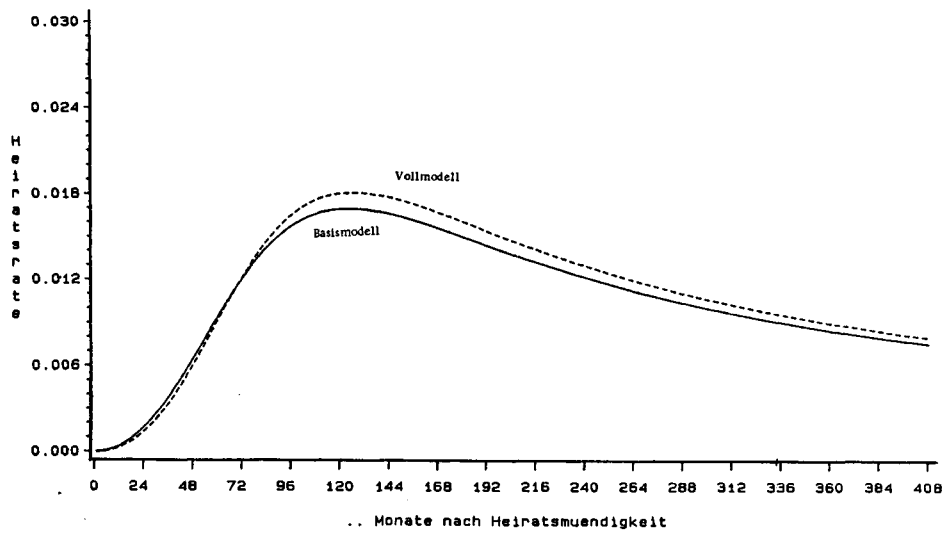
Schaubild 15
Prädizierte Heiratsraten von Männern   der Kohorte 1949-51 Basis- und Vollmodell

.. Monate nach Heiratsmuendigkeit
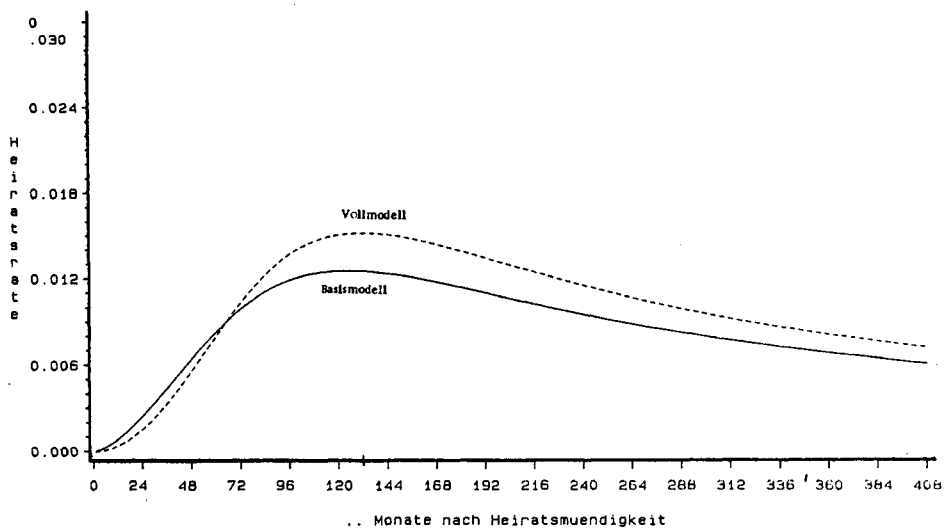
Schaubild 16
Prädizierte Heiratsraten von Männern der Jahrgänge 1929-31 und 1949-51
Basis- und Vollmodell
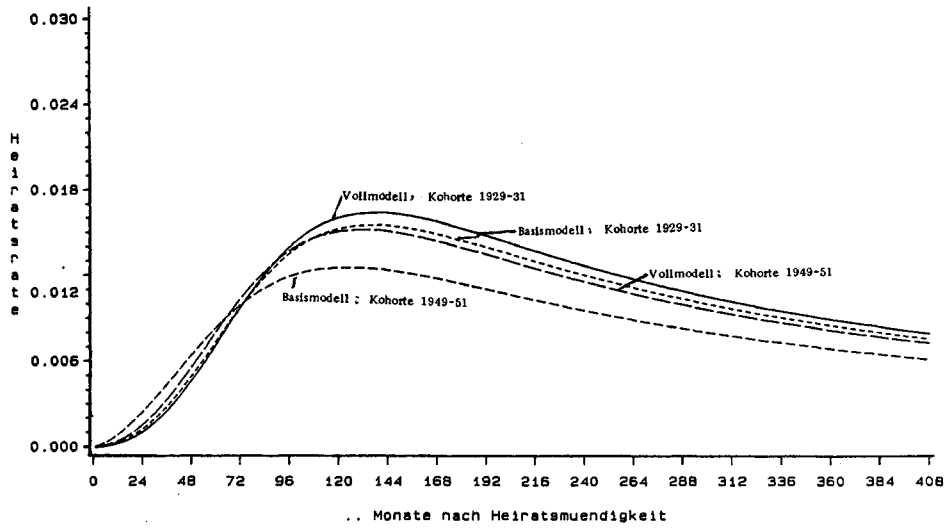
Schaubild 17
Prädizierte Heiratsraten von Frauen der Kohorte 1929-31, Basis- und Vollmodell
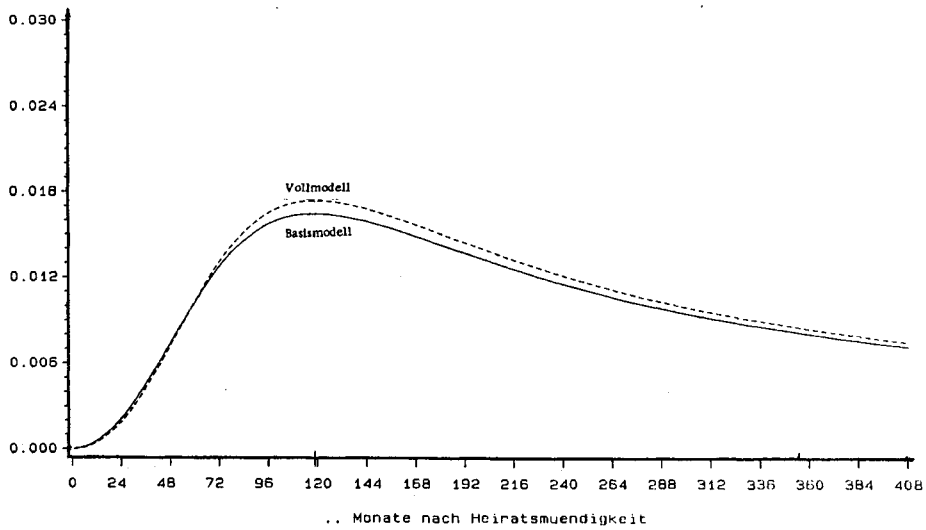
Schaubild 18
Prädizierte Heiratsraten von Frauen der Kohorte 1939-41, Basis- und Vollmodell
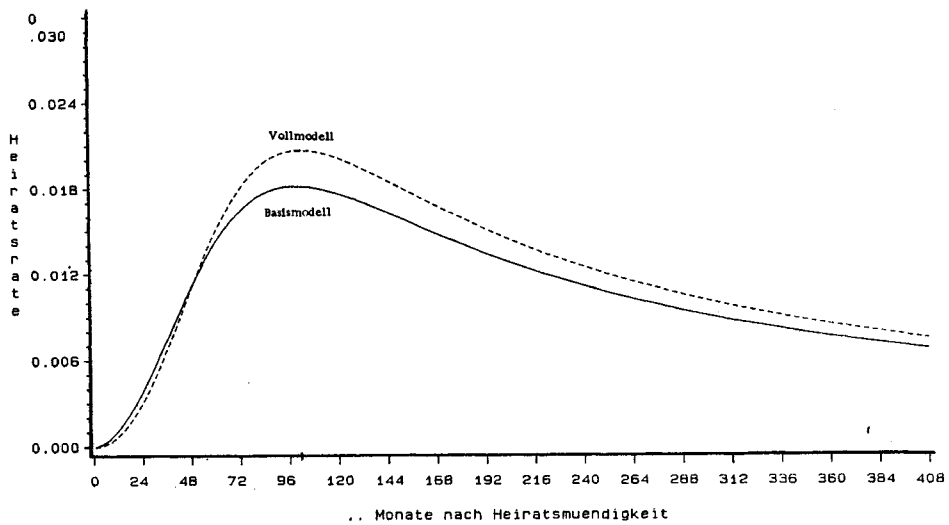
Schaubild 19
Prädizierte Heiratsraten von Frauen der Kohorte 1949-51, Basis- und Vollmodell

.. Monate nach Heiratsmuendigkeit
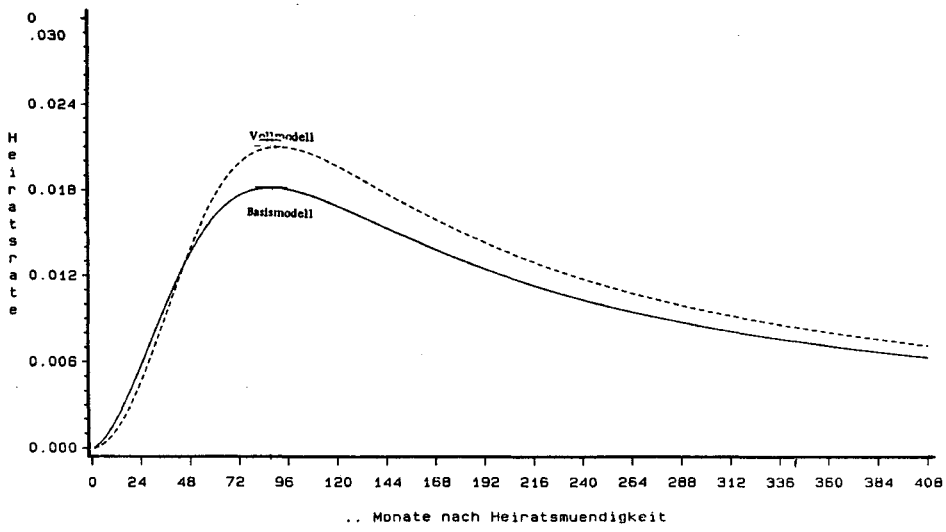
Schaubild 20
Prädizierte Heiratsraten von Männern der Kohorten 1929-31, 1939-41, 1949-51
Vollmodell

.. Monate nach Heiratsmuendigkeit
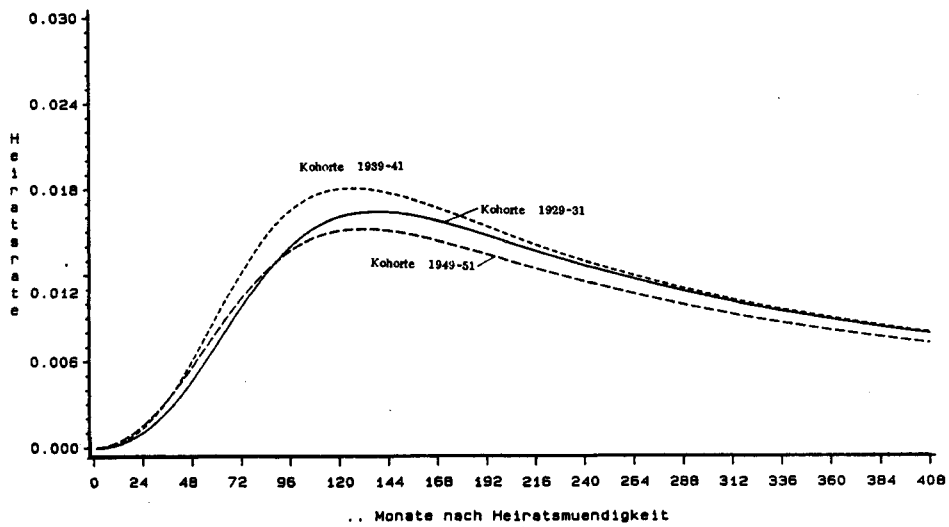
Schaubild 21
Prädizierte Heiratsraten von Männern der Kohorten 1929-31, 1939-41, 1949-51, am Maximum
der Kohorte 1929-31 zentriert, Vollmodell

.. Monate nach Heiratsmuendigkeit
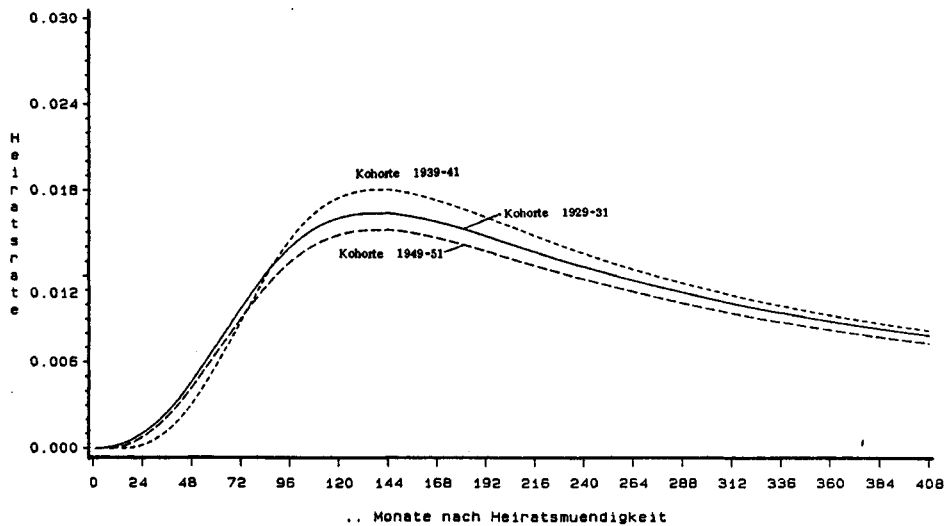
Schaubild 22
Prädizierte Heiratsraten von Frauen der Kohorten 1929-31, 1939-41 und 1949-51
am Maximum der Kohorte 1929-31 zentriert, Vollmodell

.. Monate nach Heiratsmuendigkeit
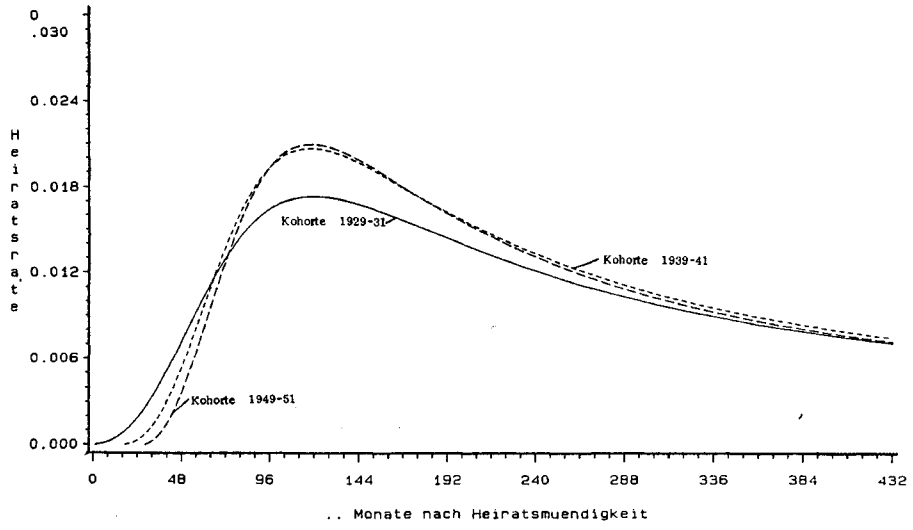
Schaubild 23
Prädizierte Heiratsraten von Frauen der Kohorten 1929-31, 1939-41, 1949-51
Vollmodell

.. Monate nach Heiratsmuendigkeit
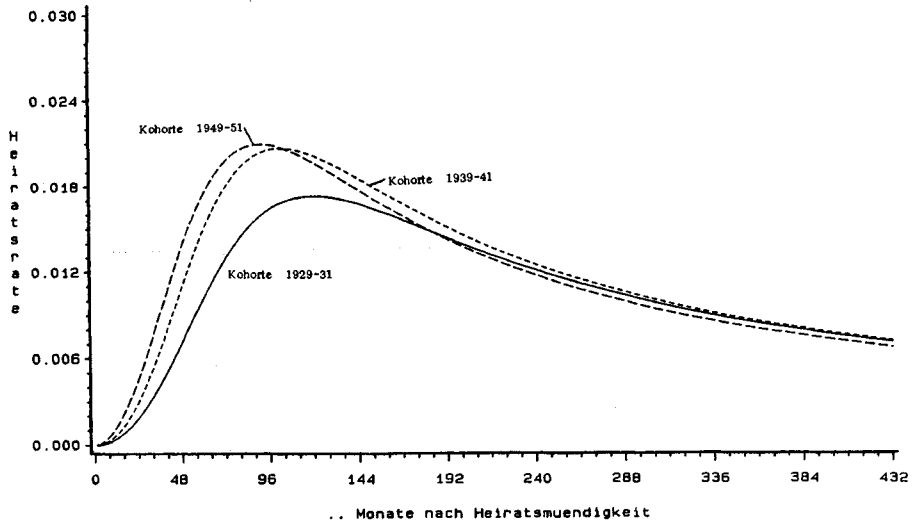
Effects of education, occupational characteristics and cohort on the "family cycle"

Andreas Diekmann

1. Introduction

Patterns of family formation and dissolution are largely influenced by socio-economic factors like education and income of spouses. A theoretical framework explaining the relations between these variables and the timing and duration of marriage is given by Becker's (1981) theory of family economics.

Many multivariate empirical studies dealing with the sociological and economic causes of divorce and age at marriage either use individual dichotomous response data (e.g. the attribute never married at some age) or aggregate data (e.g. divorce rates in different states). In the former case data are analysed by logit- or probit-models or some other dichotomous-response model (see e.g. Becker, Landes, Michael, 1977; Galler, 1979; Hogan, 1978; McDonald and Rindfuss, 1981; Michael, 1979; Waite and Spitze, 1981) while in the latter case conventional regression methods can be applied (Freiden, 1975; Preston and Richards, 1975; Sander, 1985). A third line of empirical research has recently begun to utilize event-history data on the "familiy cycle". Age at entry into marriage or marriage duration are regarded as "arrival times" which are analyzed by appropriate methods of survival analysis. This kind of analysis is usually superior because it builds on very informative event-history data, yields a model for the process, and allows for many derivations of characteristics of the process (see Hannan, Tuma, Groeneveld, 1977; Teachman, 1982; Heaton, Albrecht, Martin, 1985; Tuma and Michael, 1985; Sørensen and Sørensen, 1985; Diekmann and Mitter 1984a; Diekmann, 1986a).

In this paper marriage and divorce data from the cumulative General Social Survey 1980 to 1984 on about 9000 West German households are analyzed by techniques of event history analysis. Three aspects are stressed in this article: To pre-

sent explanations for the typical non-monotonic hazard functions of the "risk" of marriage and divorce, to evaluate the relative strength of effects of different socio-economic variables as well as cohort effects, and to estimate the effects of expansion in education on the marriage and divorce patterns.

2. Theory

While there is a long tradition of research on marriage and the family by psychologists and sociologists (Goode, 1963; Carter and Glick, 1970; Cherlin 1981) economists only recently became interested in this research area. Becker's innovative idea was to apply the apparatus of modern micro-economic theory, i.e. marginal analysis, to the process of family formation and dissolution (Becker, 1981; Becker, 1975; Becker, Michael, Landes, 1977; Michael, 1979) Starting with a sparse set of principles the theory allows derivation of many different hypotheses and integrates a lot of well known sociological findings in a common framework.

The basic assumption of the theory is that a household production function with time for household work, time for market work, market goods, prices of market goods, and wage rates as inputs is maximized under budget and time restrictions. People marry if the value of this function is higher for a common household than the utility of remaining single, i.e. if there is a gain through marriage. This gain through marriage is higher, the higher the similarity or complementarity between traits that are used jointly in production and the more dissimilar are skills of the household members which can be substituted for each other. In accordance with the theory women with high income, occupational prestige, and human capital profit less from marriages because they lose by the traditional mode of division of labor in the household. Therefore, high income women with a high degree

of education marry later or not at all and are more prone to divorce. A strong sex-specific interaction effect is expected for these varables. The theory predicts that men with high income, occupational prestige, and human capital gain much from marriage. These men marry early after finishing education and are less prone to divorce.

Of course, education increases the age of marriage for both sexes because the "risk" of marriage is low for both men and women as long as these are in the educational system (Galler, 1979). However, after finishing college or university the degree of education has a positive impact on the tendency to marry for men and a negative impact for women. Hence, the shift of age at marriage by education must be stronger for women than for men (Keeley 1977: 245).
Costs of divorce is another component influencing both age at marriage and marriage stability. The longer the expected duration of marriage the longer is the period of search for a mate if information is imperfect. Because it is assumed that catholics have higher costs of divorce, a higher age at marriage and a lower risk of divorce is expected for the catholic population other things being equal. On the other hand if people's search-period is short the likelihood of a "mismatch" increases. This might be a reason for the marriage-destabilizing-effect of early marriages.As can be seen from this short discussion Becker's theory is much more a framework than a closed deductive theory. The core of the theory is the hypothesis of utility maximization under restrictions. However, additional auxiliary hypotheses are necessary to bridge the gap between theoretical terms like "gain through marriage" and the observed socio-economic characteristics.

3. Model and Data

The aim of this study is to explore determinants of changes between the discrete states "unmarried", "first marriage" and "(first)divorce". A continous-time-discrete-state semi-Markov-process seems to be an appropriate model because events may happen at any point in time and the risk of change is allowed to be duration-dependent. Marriages are not only dissolved by divorce but also by the "competing risk" of death of one of the spouses (Figure 1).
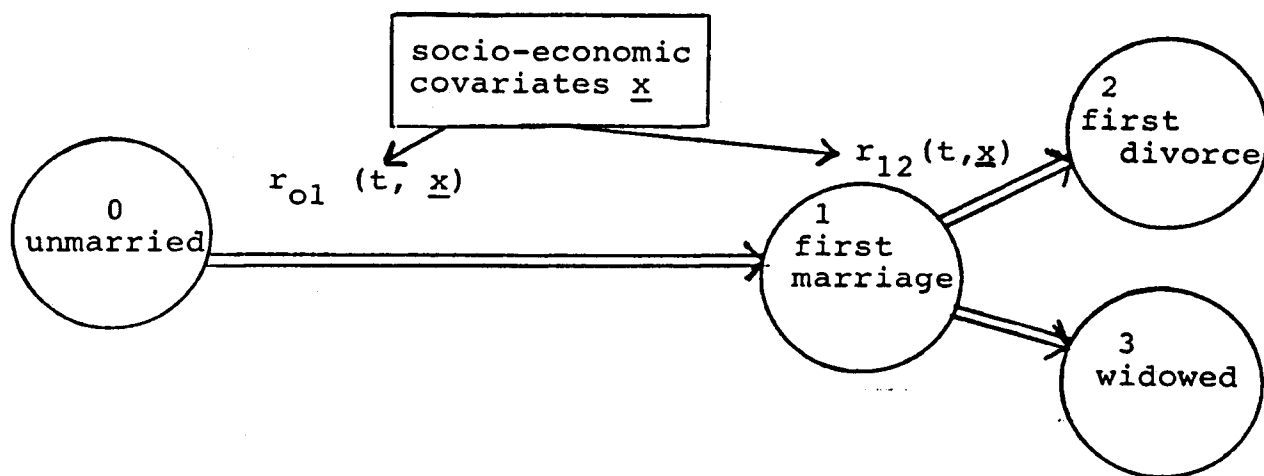
Figure 1   A Model for the change of family states

Observed duration times until the dissolution of a marriage by death are treated as censored data. As usual in event-history analysis (Tuma and Hannan, 1984; Diekmann and Mitter, 1984b) a hazardrate or "risk"-equation incorporating the socio-economic covariates is estimated by methods of Maximum-Likelihood or Partial-Likelihood. The hazardrate r_{o1} (t, $\underline{x}$) can be interpreted as the "risk" or "instantaneous" likelihood of a first marriage, while the rate r_{12} (t, $\underline{x}$) is the risk of divorce (Figure 1). Formally the rate is defined as:

$$(1) \quad r_{jk} (t, \underline{x}) = \lim_{\Delta t \to 0} \frac{Prob [t + \Delta t > T_j > t | T_j > t]}{\Delta t} = \frac{f_j(t)}{G_j(t)},$$

where $\underline{x}$ is a vector of covariates, T_j is the random variable "waiting time" until a change to state k, $f_j(t)$ is the probability density distribution of waiting times referring to state j and $G_j(t)$ is the probability of survival (survival function) in state j.

As mentioned in the introduction, parameters are estimated on data of the cumulative General Social Survey 1980 to 1984 (Allbus 80-84). The analysis of "unmarried episodes" utilizes only the 1982 and 1984 sample (about 6000 households) because there was the danger of a systematic bias by a different operationalization of "age" in the 1980 survey. Estimation of effects on divorce likelihood refers to the subset of first marriages from all three surveys. The sample size is further reduced by missing data, particularly if income is included in the equations, or by the necessity to draw a random sample in case of Cox-regression because of limited computer space.

4. Explanations of Non-monotonic Risk Functions

Life-table-estimates of the hazardrate for first marriage follow a non-monotonic, bell-shaped pattern (Figure 2). This "marriage bell" corresponds with a typical S-shaped cumulative distribution of age at marriage. Survival-functions for women and men are of the same shape, delayed by two to three years with a crossover point at age 41 (Figure 3). The proportion of never married women at age 50 (6.9 %) is a little higher than the proportion of never married men (5.3 %).

S-shaped cumulative distributions and non-monotonic bell-shaped hazardrate functions are reported by many authors studying entry into marriage (e.g. Sørensen and Sørensen, 1985; Espenshade, 1982). Hence, the "marriage bell" seems to be a cross-cultural invariant law. How can this regularity be explained?
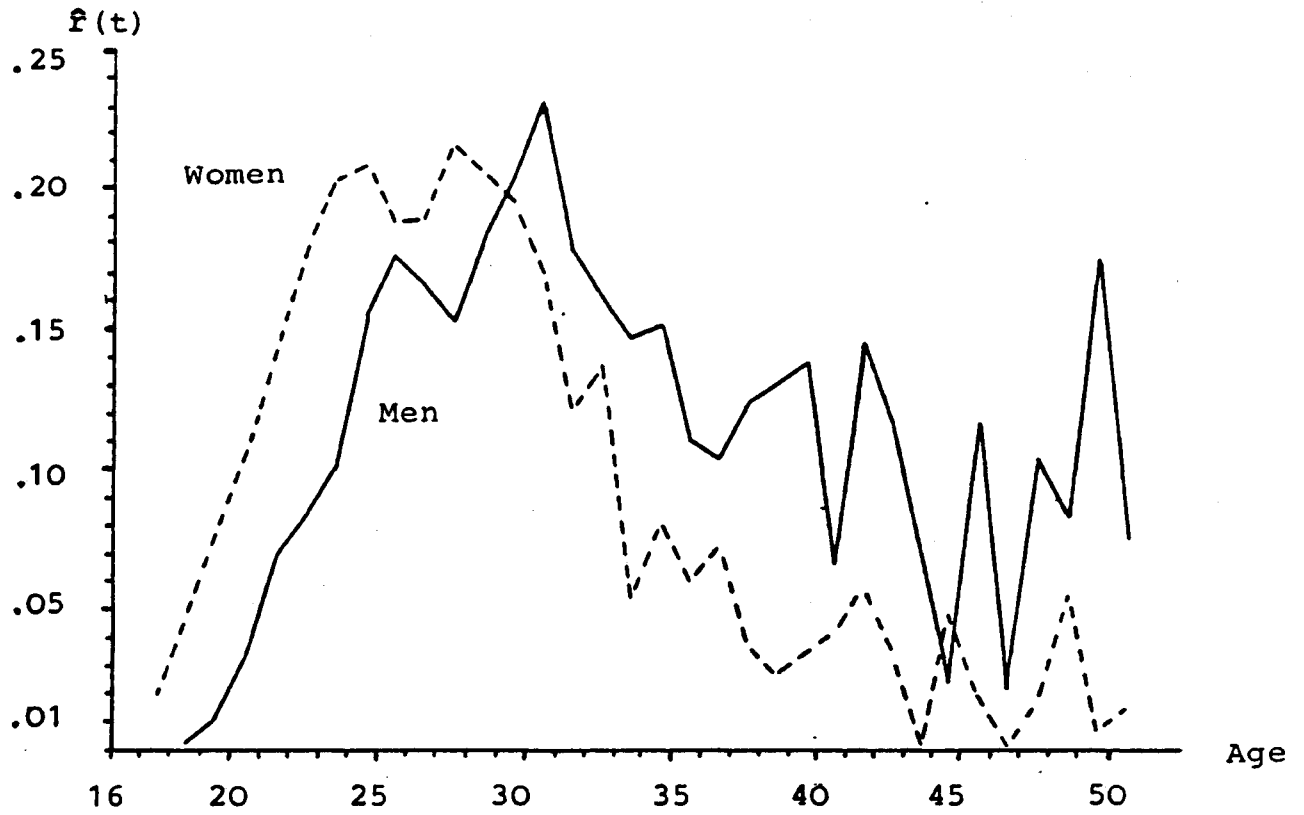
Figure 2 Life table hazardrate estimates for entry into first marriage by sex

G(t), proportion unmarried
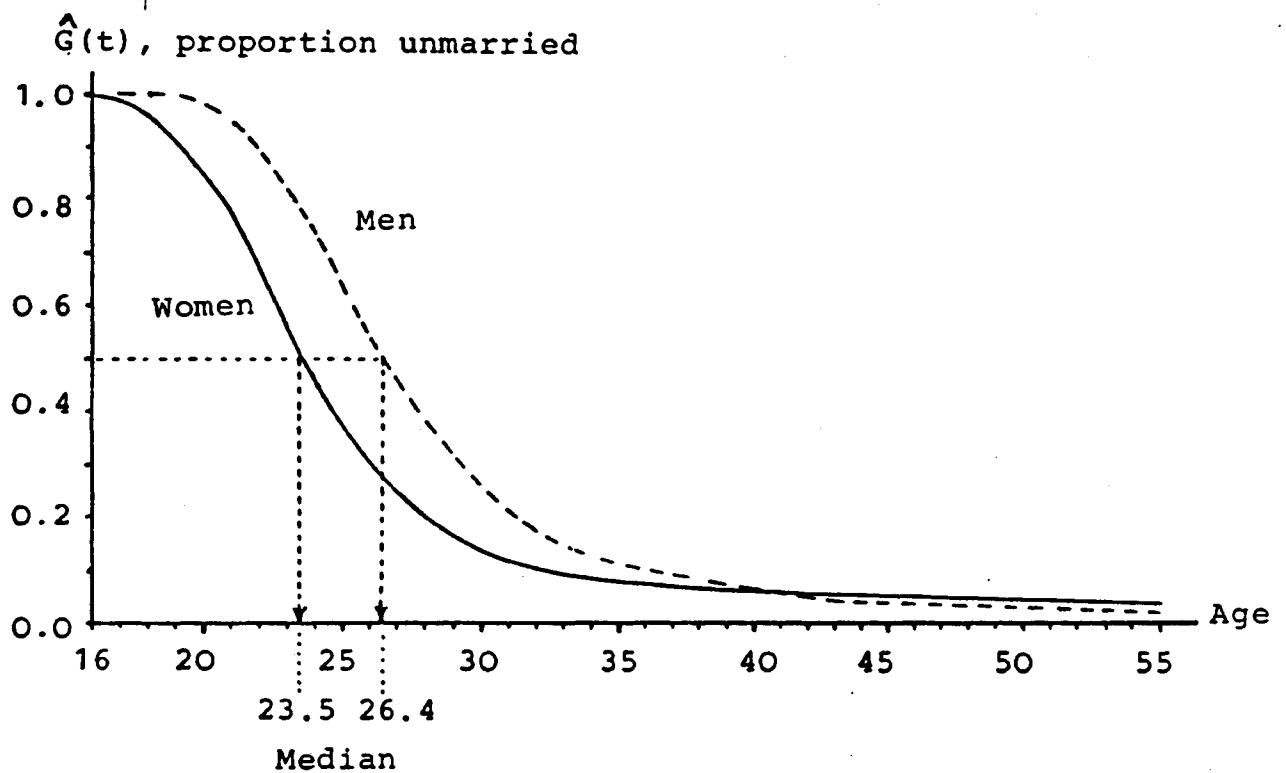
Figure 3 Life table survival estimates by sex

Hernes (1972) successfully applied the following model, similar to diffusion-of-innovation-theory, to the cumulative distributions of age at marriage of US-birth-cohorts:

$$(1) \quad \frac{dP(t)}{dt} = s(t) \ P(t) \ (1-P(t))$$

with

$$(2) \quad s(t) = m \cdot exp \ (-c(t-1)), \quad t \geq 1.$$

Here P(t) is the proportion first married at age $t + t_o$ (t_o is the earliest possible age of marriage which was by law 16 for women and 18 for men in Germany), s(t) is a decreasing function of time, and m,c are constants estimated by empirical data.

In this model the increase in proportion first married per
unit of time and per candidate for a marriage, i.e.
dP(t)/(dt(1-P(t))) is the product of two terms: The increa-
sing social pressure P(t) inducing imitation or "infection",
and a decreasing chance of marriage s(t) dependent on age.
Hernes (1972) did not deal with hazardrate models but de-
rived from differential equation (1) and (2) the solution
for the cumulative distribution P(t).

However, there is a simple correspondence between diffusion-
models of the above type and hazardrate models. Because

$$G(t) = 1 - P(t) \quad \text{and} \quad f(t) = dP(t)/dt$$
it follows from (1):

$$(3) \quad r(t) = \frac{f(t)}{G(t)} = s(t) \cdot P(t).$$

It can be shown that hazardrate (3) is in fact non-monotonic.

As an alternative to Hernes (1972) choice of s(t) there is the
simple function:

$$(4) \quad s(t) = \frac{p}{t}$$

In this case it follows from (1), (4), and (3):

$$(5) \quad P(t) = \frac{(\lambda t)^p}{1+(\lambda t)^p}$$

(6) $r(t) = \dfrac{\lambda p \ (\lambda t)^{p-1}}{1 + \ (\lambda t)^p}$

where parameter λ results from the constant of integration. (5) is the log-logistic hazardrate model with non-monotonic hazardfunction (6) in case of $p > 1$. The widely used log-logistic model is, therefore, not only a descriptive model but can be derived from differential equation (1) with specification $s(t) = p/t$. [1]

Both variants of the diffusion model (1), the Hernes-model and the log-logistic model, were tested against the data. The three parameters of the Hernes model (a third parameter results from the integration constant) are estimated from the life-table values of the survival-function by a method suggested by Prescott (1922, see also Hernes, 1972). Parameters λ and p of the log-logistic model are estimated by the principle of Maximum-Likelihood using GLIM. [2] In addition to these two models the sickle function, applied by Diekmann and Mitter (1984a) to divorce data was tested as a third candidate.

Results are shown in Figure 4. The Hernes model as well as the log-logistic model yields a quite good fit of the data. [3]

$\hat{G}$ (t), proportion unmarried

a) Women

$\hat{G}$(t), proportion unmarried
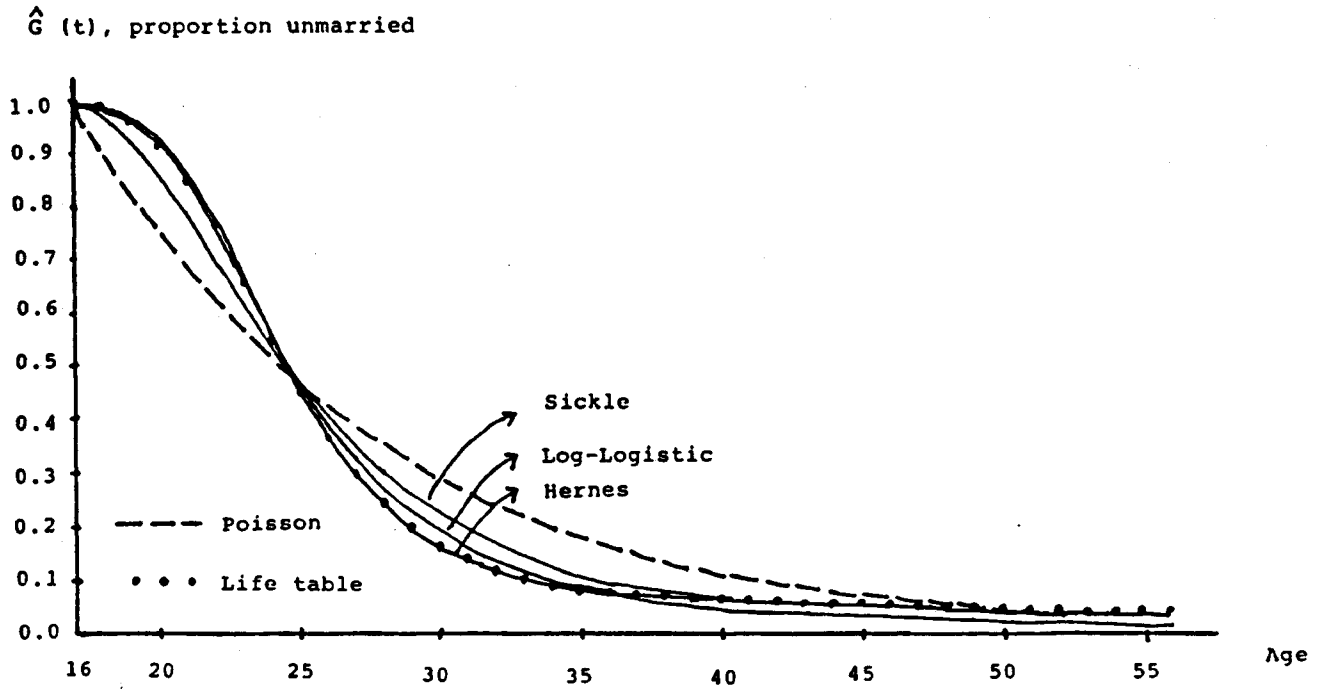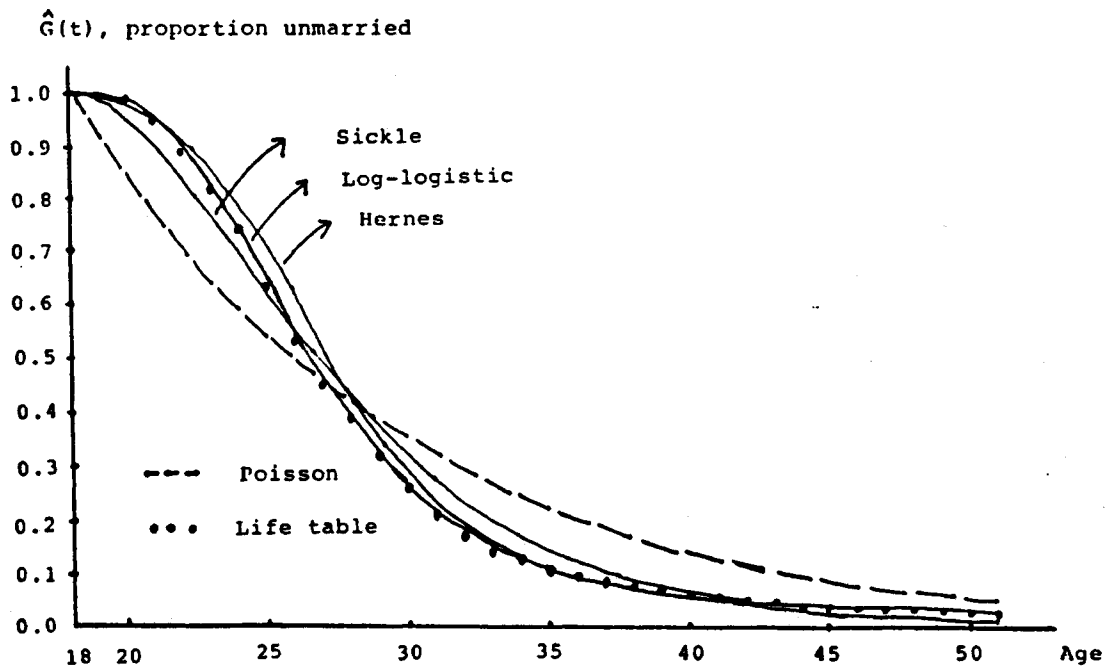
b) Men

Figure 4 Application of three hazardrate models to age-at marriage-data

The two parametric log-logistic model seems to be a "sparse" and appropriate model for the entry-into-marriage process, although this model lacks the desirable property of a "defective" distribution. The Hernes model and the sickle model, on the other hand, allow for a proportion of never married people. However, this is not a serious disadvantage of the log-logistic model because there is only a small proportion of never married people.

Things are different for divorce data. The proportion of never divorced couples is very high (between 70% and 90%) and therefore, the log-logistic model highly inappropriate. As was shown by Diekmann and Mitter (1984 a) the sickle model yields a very good fit for cohort data of marriage duration. The risk of divorce described by the sickle function is non-monotonic as well. A model explaining the non-monotonic risk-function of divorce is presented in Diekmann and Mitter (1984a) but there is also another explanation building on search theory and heterogeneity (Becker, 1975: 338). The search of a mate under imperfect information is based on "search traits" and "experience traits". There is fuller information on search traits like education but less information on experience traits. It follows that some proportion of the married population made "mistakes", which are not detected before marriage. These marriages have an increasing risk of divorce. On the other hand the risk of divorce declines with an increase in marriage specific investments. Both processes, the increasing divorce risk of marriages with "mistakes" and the declining divorce risk of marriages with specific investments may produce a non-monotonic divorce risk over the married population.

5. Effects of Socio-Economic Covariates
5.1. Age at Marriage

Life table analysis reveals that there are strong cohort effects in respect to age at marriage (Table 1). For both men

and women a U-shaped pattern can be observed. If censored data are ignored, i.e. if the median age at marriage is calculated for the married group only, a strong bias depending on the degree of censoring is introduced (Table 1). Neglecting censored data would lead to the erroneous conclusion of a continuously declining age at marriage.

For multivariate analysis of covariate effects on the hazard-rate of marriage (and divorce risk in 5.2.) Cox' proportional hazards model is applied, whereby parameters are estimated by the partial likelihood method (Kalbfleisch and Prentice, 1980). This semi-parametric model allows for unspecified time dependence and, therefore, also for non-monotonic duration dependence as discussed in part 4. The hazardrate-function is specified as follows:

$$(7) \quad r(t) = \lambda_0(t) \; \alpha_1^{x_1} \; \alpha_2^{x_2} \; ... \alpha_m^{x_m}$$

with $\lambda_0(t)$ an unknown function of time, covariates x_1 to x_m, and parameters α estimated by partial likelihood. α-values (Table 3) give information about percentage effects on rate $r(t)$. The percentage effect is $(\alpha - 1) \cdot 100$. An estimate of .86 (catholic women in Table 3), for example, means that there is a 14 per cent lower hazardrate (and, therefore, a later age at marriage) for catholic women in comparison to non-catholic women other things being equal.
The U-shaped cohort pattern does not disappear if other socio-economic variables are incorporated in a multivariate analysis. Table 3 shows an inverse U-shaped risk pattern: There is a relatively low hazardrate for respondents born before 1926 (reference category BC5) which increases until cohort 1946-55 for women and cohort 1936-45 for men and then decreases. This pattern corresponds to the trend in Table 1.

Birth cohorts	Women	Men	Median age for married men +	% censored
1956-66	24.4	27.9	22.5	82.6
1946-55	22.2	25.8	24.6	24.2
1936-45	22.9	25.2	25.1	4.0
1926-35	23.6	26.0	25.9	2.0
before 1926	24.7	28.2	28.1	1.9

+ Life table analysis neglecting censored data

Table 1 Cohort differences in median age at marriage by sex

Education	Women		Men	
	Median	Proportion never married	Median	Proportion never married
9 years school with certification*	23.0	.032	25.7	.020
10 years school**	24.1	.064	26.1	.027
13 years school with university admission***	26.7	.149	29.2	.038

* "Hauptschulabschluß"
** "Realschule, Mittlere Reife"
*** "Abitur, Hochschulreife"

Table 2 Median age at marriage and proportion never married at age 50 by sex and education

			Women	Men
			1.68*	1.32
	BC1	1956-66	(4.43)	(1.93)
			1.89*	1.44*
Birth	BC2	1946-55	(7.74)	(3.99)
			1.73*	1.49*
Cohorts	BC3	1936-45	(6.91)	(4.68)
			1.44*	1.28*
	BC4	1926-35	(4.46)	(2.95)
Religion (catholic versus other) (Rel)			.86* (2.75)	1.11 (1.62)
Occupational prestige father (OccFa)			.9974 (.90)	.9908* (2.92)
Edu- ca- tion	E1	8 to 9 years school without certification	1.22 (.60)	1.01 (.02)
	E3	10 years school	.73* (4.30)	.91 (1.20)
	E4	12 years school equivalent	.52* (3.22)	.98 (.17)
	E5	13 years school, university admission	.44* (6.40)	.66* (4.21)
City size (citmed)			1.19* (2.64)	1.10 (1.39)
Chi^2 (df=11)			139.95	70.90
number of cases N			1580	1478
% censored			17.6%	25.6%

α-effects (α=expβ), β-coefficient divided by standard error in brackets. * significant for p=.05. Reference category: 9 years school with certification ("Hauptschulabschluß") E2, non-catholic, birth-cohort before 1926 (BC5), City size below 2000 and greater or equal 500.000 (citmed=1 denotes medium city size). Occupational prestige measured on the Treiman scale. E1: without "Hauptschulabschluß", E2: "Hauptschulabschluß, E3: "Mittlere Reife", E4: "Fachhochschulreife", E5: "Abitur, Hochschulreife". Data are 70% random sample from "ALLBUS" 1982 and 1984.

Table 3 Partial-Likelihood estimation of socio-economic effects on hazardrate for entry into marriage

The search theoretical hypothesis concerning the effect of religion can be confirmed for women but not for men. Educational effects are very well in line with the theory. There is a monotonically declining hazardrate with educational degree for women and, also, a significant effect of higher education for men. As expected education effects are stronger for women than for men. In addition, bivariate life table analysis yields an estimate of never married women with higher education at age 50 of 15% in comparison to 4% for men (Table 2). Father's occupational prestige reduces the hazardrate at least for men while this effect is not significant in the women's analysis. The covariate "city size" was included in the model with the coding reported in Table 3 because lifetable-analysis and multivariate analysis with several categories of size suggested that there is a low rate for small villages as well as for very large cities. This U-effect of size is significant for women but should be interpreted with caution because measurement of city size refers to time at interview.

Also, the results reported in Table 4 require some qualifications. Measurement of occupational prestige and income was at time of interview. However, the coefficients may be interpreted as effects on the hazardrate for entry into marriage if one assumes that the two variables are highly correlated with income and prestige before marriage. Number of children of respondents of age 40 and older should serve as an indicator for the preference for children before marriage. With these qualifications in mind reported coefficients at least do not refute the theory, except for the negative effect of men's occupational prestige which is, however, not significant. A significant income-sex interaction effect is

	Women	Men
Occupational prestige I (Occ)	.9915* (2.54)	.9960 (1.23)
Income II (Inc)	.9846* (2.38)	1.0106* (4.83)
Number of III children	1.22* (11.62)	1.20* (8.23)

I, II, III separate PL-estimations with controls: birth cohorts, education, OccFa, Rel, Citmed (see Table 1 for definition of terms). Income is net monthly income measured in units of DM 100.- at time of interview. III is estimated on the subsample of men and women of age 40 and older.

Table 4 Partial-likelihood estimation of coefficients for occupational prestige, income, and number of children

observed and a negative effect of women's occupational prestige as predicted. A high hazardrate and, therefore, low age of marriage is found for respondents with a higher number of children. However, especially the negative income effect for women and the positive income effect for men can also be explained by an alternative mechanism. An increase in men's income and a decrease of women's income after marriage as usually observed in income studies might be responsible for the results reported in Table 4. Retrospective data on covariates or longitudinal studies are necessary in order to scrutinize effects of occupational careers on the family dynamics.

A comparison of different model specifications shows that parameter estimations are rather robust despite misspecifi-

cation of models (2) to (4) in light of the non-monotonic
hazardrate pattern in Figure 2. Significant effects of the
correctly specified proportional hazards model do not devia-
te very much from corresponding estimates of the alternative
models (Table 5).[4] Of course these results should not be
misunderstood as providing a guarantee of always receiving
good estimates from wrong models.

	(1) Proportional Hazards	(2) Exponential (Constant rate)	(3) Gompertz	(4) Gamma
BC1	1.014	.7384*	.7081*	.7334*
BC2	1.808*	1.784*	1.726*	1.778*
BC3	1.667*	1.740*	1.688*	1.732*
BC4	1.416*	1.372*	1.351*	1.372*
Rel	.9582	.9420	.9473	.9434
OccFa	.9991	.9950	.9955	.9951
E1	1.007	.7759	.7923	.7780
E3	.7659*	.8043*	.8066*	.8033*
E4	.3768*	.5014	.5001	.4992
E5	.4914*	.5481*	.5546*	.5468*
Income	.9879*	.9904	.9907	.9905
c_1	-	.1037*	.1072*	.1043*
c_2	-	-	-.004769	.008511
Chi^2	88.29	108.54	110.11	108.60
df	11	11	12	12

Women. N=1007. 22.9% censored data. See Table 1 and Table 2 for definition of variables.
Estimation of Models (2) to (4) by RATE (Tuma 1979).
Models:

(1) $\quad r(t) = \lambda_o(t)\alpha_1^{x1}\,\alpha_2^{x2}\,\ldots$

(2) $\quad r \quad = c_1\,\alpha_1^{x1}\,\alpha_2^{x2}\ldots$

(3) $\quad r(t) = c_1\alpha_1^{x1}\,\alpha_2^{x2}\,\ldots\exp(c_2 t)$

(4) $\quad r \quad = \varepsilon c_1\alpha_1^{x1}\,\alpha_2^{x2}\,\ldots$

where ε is a gamma distributed error term with mean 1 and variance c_2.

Table 5 A comparison of four specifications of the hazardrate for entry into marriage

5.2. Divorce Risk

As with age at marriage data a U-shaped cohort effect is observed. Cohorts are defined by time of marriage and not by birth. Divorce risks were high among pre-war marriages possibly caused by consequences of war (reference category in Table 6). Risks decreased for marriage cohorts until mid sixties and increased afterwards to a higher than pre-war level.

As expected by theoretical reasons divorce is more likely for women with higher education. In addition there is no significant effect of men's education. Catholic religion reduces the risk while city size has a positive impact on divorce. However, because measurement was at time of interview it may also be a consequence of divorce. Particularly divorced women may move to larger cities because of lower discrimination and better job opportunities. In line with the hypotheses stated in part 2 age at marriage influences strongly the risk of divorce. A one year delay in marriage reduces the divorce risk by about 7 to 8 per cent. No effect of occupational variables, neither father's nor respondent's occupational prestige, was observed for women and men. This result clearly contradicts the theory.

Empirical evidence reported herein is somewhat ambiguous for the hypotheses derived from the economics of family theory. While educational effects are in line with the theory this is not true for occupational variables in respect of the risk of divorce.

			Women	Men
Mar- ria- ge co- horts	MC1	1975-84	1.64 (1.35)	5.38* (4.41)
	MC2	1965-74	1.55* (2.13)	2.28* (2.87)
	MC3	1955-64	.80 (1.06)	.80 (0.73)
	MC4	1945-54	.60* (2.18)	.85 (0.53)
Edu- ca- tion	E3	10 years school	1.41 (1.87)	1.03 (.12)
	E5	13 years school,uni- versity ad- mission	2.60* (3.19)	1.46 (1.14)
	Religion (Rel) (catholic ver- sus other)		.74* (2.04)	.69 (1.88)
	City size (Citlarge)		1.74* (3.34)	1.52 (1.96)
Age at marriage			.93* (3.57)	.92* (3.46)
Occupational prestige of father (OccFa)			.9893 (1.40)	.9994 (.06)
Occupational prestige (Occ)			.9958 (.54)	.9943 (.57)
Chi^2 (df=11)			68.56	77.54
N			2267	2269
% censored			91.4 %	94.6 %

First marriages "Allbus" 1980-1984, E1 and E4-persons ex-
cluded because low numbers of cases and events in these ca-
tegories. Citlarge = 1 for Citysize of 500000 and higher.
Reference category: 9 years school with certification
("Hauptschulabschluß") E2, marriage before 1945 (MC5), non-
catholic, and Citysize less than 500000.

Table 6   Partial-likelihood estimation of socio-economic
           effects on divorce risk

Cohort effects are "genuine" and do not disappear after in-corporation of socio-economic variables considered in this analysis (see Sørensen and Sørensen, 1985, for similar re-sults). Finally it should be kept in mind that theory-confirming results like educational effects are also explain-able by alternative hypotheses, for instance a subjective emancipation effect of higher education (see Galler, 1979).

6. Demographic Effects of Expansion in Education

Educational expansion in West-Germany in the mid-sixties and seventies is characterized by a general growth of participa-tion in higher education, a more than proportional increase or participation rates in higher education of women, and, therefore, a tendency towards equalization or sex-specific educational chances in respect to the quantitative amount of education. This trend is mirrored by the cohort specific dis-tributions over educational degrees for women and men shown in Table 7.

From the foregoing discussion it should be clear that this process has an impact on demographic patterns of marriage. But what is the quantitative amount of the demographic con-sequences of the change in educational participation rates?

	Women		Men	
	Birth cohorts		Birth cohorts	
	1936-45	1956-63	1936-45	1956-63
E1	.4	.4	.1	1.7
E2	68.3	34.1	58.1	38.9
E3	22.9	35.6	20.9	25.9
E4	1.9	5.5	6.0	5.5
E5	6.5	24.4	14.9	28.1

Allbus 1982 and 1984. Respondents older than 20. E1 for cohort 1956-63 is inflated because there might be some respondents still in high school at time of interview. See Table 1 for definition of categories.

Table 7 Expansion in education by sex and birth cohorts

	Women	Men
Median life Table	23.52	26.41
Predicted median by Proportional Hazard model with mean values of covariates	23.38	26.29
Predicted median with educational distribution of cohort 1936-45	23.29	26.23
Predicted median with educational distribution of cohort 1956-63	24.10	26.47
Shift of age at marriage by educational expansion	.81	.24

Table 8 Effect of expansion in education on median age at marriage

$\hat{G}$(t), proportion unmarried
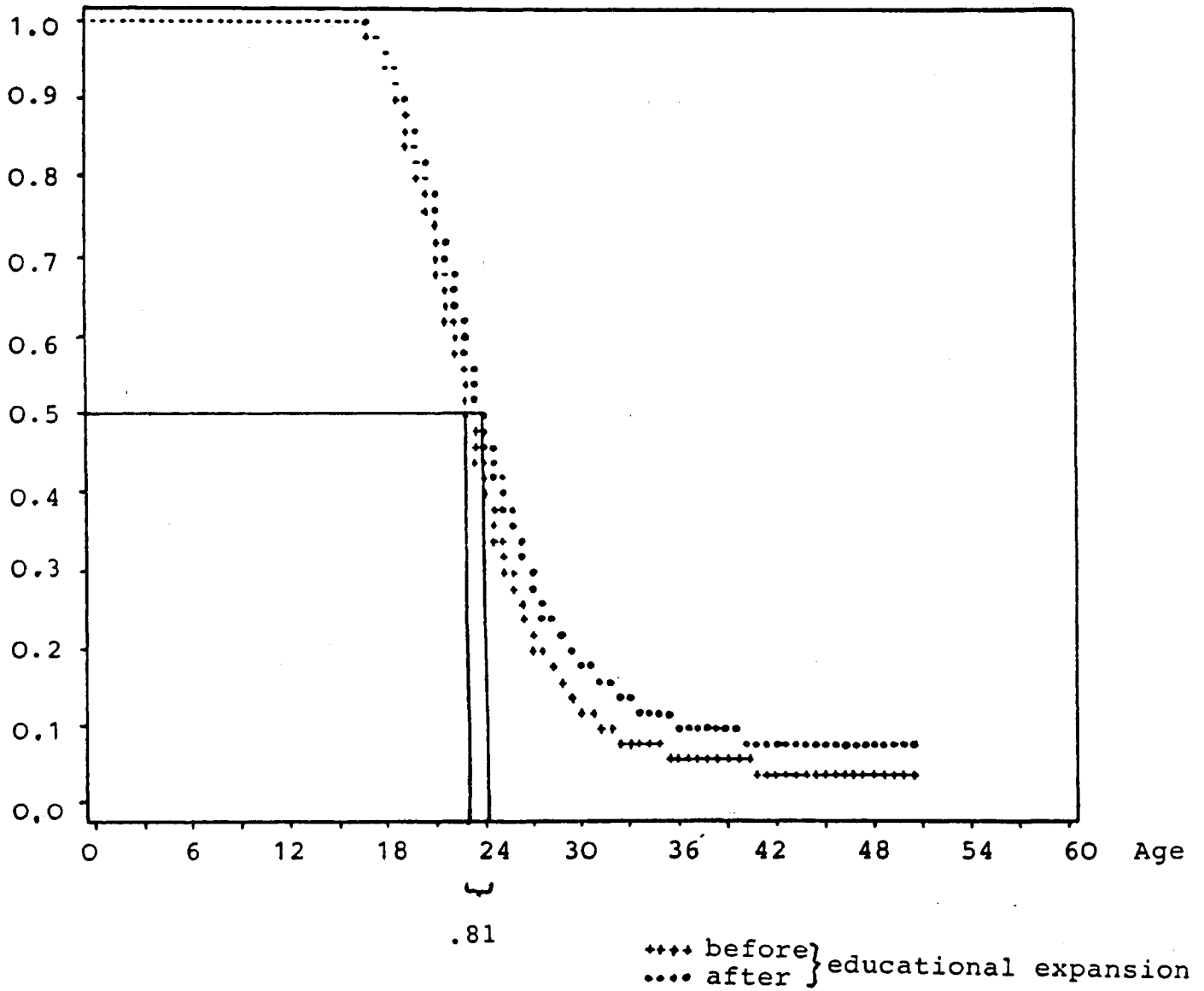
Figure 5 Effect of educational expansion on the unmarried proportions by age (survival function) for women

On the basis of proportional hazards model discussed in sec-
tion 5 an answer to the question of the "pure" effect of
educational expansion on the age at marriage can be given.
This is done by a simulation whereby the educational distri-
butions of the pre-expansion conort and post-expansion
cohort in Table 7 are incorporated in the model respectively
and all other covariates are replaced by their mean values.
This procedure requires the assumption that there is no
structural change, i.e. no snift of parameters during the
period of educational expansion. The resulting survival
functions predicted by the model give information about the
shift of age at marriage. The predicted medians are
contained in Table 8, and Figure 5 depicts the
survival-functions for women. The median shift induced by
educational expansion is about three and a half times nigher
for women (0.81 years) than for men (0.24 years). Hence, the
educational expansion also contributes to an equalization of
differences in marriage age for women and men. However,
despite tnis equalization effect the gross difference in age
at marriage of spouses increased in recent years. The larger
impact for women is not surprising in light of tne
estimations reported in section 5.1 and the change of educa-
tional distributions contained in Table 7. There is a greater
effect of education for women on the age at marriage and, in
addition, there is a greater increase in participation rates
for women in higher education than for men.

The calculated demographic consequences of the educational
expansion seem at first sight to be relatively low. However,
one should be reminded of the fact that the whole process of
educational expansion leads to an increase of average dura-
tion in the eaucational system of not more than about one
year.

Footnotes

+ I would like to thank Iain Paterson for helpful comments.

1) See Sorensen and Sorensen (1985) for an application of this model to cohort shifts in age at marriage.

2) I would like to thank Gilg Seeber who estimated the parameters of the log-logistic model with a GLIM-Macro.

3) The good fit of diffusion models does not exclude alternative explanations of the risk pattern. Assumptions of heterogeneity, search-models, and a combination of search theory, latent states and heterogeneity (Coale and McNeil, 1972) offer rival explanations of the observed patterns (see Diekmann, 1986b for a more extensive discussion).

4) Tests of the proportional hazards model are discussed in Diekmann (1986b). Kaplan-Meier estimates of the survival function fit quite well with the model prediction for mean values of covariates.

References:

BECKER, G.S. (1975), "A theory of marriage", in: T.W. Schultz, ed., Economics of the Family. Marriage, Children, and Human Capital, Chicago and London: Univ. of Chicago Press, 299-344

BECKER, G.S. (1981), A Treatise on the Family, Cambridge: Harvard Univ. Press

BECKER, G.S., LANDES, E.M., and MICHAEL, R.T. (1977), "An economic analysis of marital instability", Journal of Political Economy, 85: 1141-1187

CARTER, H. and GLICK, P.C. (1970), Marriage and Divorce: A Social and Economic Study, Cambridge, Mass.: Harvard University Press

CHERLIN, A.J. (1981), Marriage, Divorce, Remarriage, Cambridge, Mass.: Harvard University Press

COALE, A.J. and Mc NEIL, D.R. (1972), "The distribution by age of the frequency of first marriage in a female cohort", Journal of the American Statistical Association, 67: 743-749

DIEKMANN, A. and MITTER, P. (1984a), "A comparison of the 'sickle function' with alternative stochastic models of divorce rates", in: A. Diekmann and P. Mitter, eds., Stochastic Modelling of Social Processes, Orlando: Acad. Press, 123-153

DIEKMANN, A. and Mitter, P. (1984b), Methoden zur Analyse von Zeitverläufen, Stuttgart: Teubner

DIEKMANN, A. (1986a), "Verlaufsdatenanalyse und Lebensverläufe", in: W. Voges, ed., Methoden der Biographie- und Lebenslaufforschung, Opladen: Westdt. Verlag

DIEKMANN, A. (1986b), "Determinanten des Heiratsalters und Scheidungsrisikos. Eine Anwendung von Modellen und Methoden der Verlaufsdatenanalyse", Ms., Institute for Sociology, University of Munich

ESPENSHADE, Th.J. (1982), "Marriage, divorce, and remarriage from retrospective data: a multiregional approach", Ms., International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

FREIDEN, A. (1975), "The U.S. mariage market", in: T.W. Schultz, ed., <u>Economics of the Family. Marriage, Children, and Human Capital</u>, Chicago and London: Univ. of Chicago Press, 352-371

GALLER, H.P. (1979), "Schulische Bildung und Heiratsverhalten", <u>Zeitschrift für Bevölkerungsforschung</u>, 5: 199-213

GOODE, W. (1963), <u>World Revolution and Family Patterns</u>, New York: Free Press

HANNAN, M.T.,TUMA, N.B., and GROENEVELD, L.P. (1977), "Income and marital events: Evidence from an income-maintenance experiment", <u>American Journal of Sociology</u>, 82: 1186-1211

HEATON, T.B., ALBRECHT, S.L., and MARTIN, Th.K. (1985), "The timing of divorce", <u>Journal of Marriage and the Family</u>, 47: 631-639

HERNES, G. (1972), "The process of entry into first marriage", <u>American Sociological Review</u>, 37: 173-182

HOGAN, D.P. (1978) "The effects of demographic factors, family background, and early job achievement on age at marriage", <u>Demography</u>, 15: 161-175

KALBFLEISCH, J.D. and PRENTICE R.L. (1980), <u>The Statistical Analysis of Failure Time Data</u>, New York: Wiley

KEELEY, M.C. (1977), "The economics of family formation", <u>Economic Inquiry</u>, 15:238-250

MacDONALD, M.M. and RINDFUSS, R.R. (1981), "Earnings, relative income, and family formation", <u>Demography</u>, 18: 123-136

MICHAEL, R.M. (1979), "Determinants of divorce", in: L.L. Lévy-Garaa, ed., <u>Sociological Economics</u>, London: Sage, 223-254

PRESCOTT, R.B. (1922), "Law of growth in forecasting demand", <u>Journal of the American Statistical Association</u>, 18: 471-479

PRESTON, S.H. and RICHARDS, A.Th. (1975), "The influence of women's work opportunities on marriage rates", <u>Demography</u>, 12: 209-222

SANDER, W. (1985), "Women, work, and divorce", <u>The American Economic Review</u>, 75: 519-523

SØRENSEN, An. and SØRENSEN Aa. (1985), "An event history analysis of the process of entry into first marriage," Ms., University of Wisconsin-Madison and Harvard University

TEACHMAN, J.D. (1982), "Methodological issues in the analysis of family formation and dissolution", Journal of Marriage and the Family, 44: 1037-1053

TUMA, N.B. (1979), "Invoking Rate", working paper ot "Zentrum für Umfragen, Methoden und Analysen" (ZUMA), Mannheim

TUMA, N.B. and HANNAN, M.T. (1984), Social Dynamics. Models and Methods, Orlando: Acad. Press

TUMA, N.B. and MICHAEL, R.T. (1985) "A comparison of statistical models for life course analysis with an application to first marriage", Ms., Stanford University, Calif.

WAITE, L.J. and SPITZE, G. (1981), "Young women's transition to marriage", Demography 18: 681-194

Marriage rates for women in the U.S.: Some exploratory analyses and methods

Larry L. Wu

Issues in Smoothing Empirical Hazard Rates

Social scientists often rely on a few parametric models when conducting event history analyses. But we often lack a sound theoretical basis for choosing between parametric models. In such circumstances, it would be useful to obtain an estimator for the hazard rate that makes minimal assumptions about the distribution of event times. This chapter describes such an estimator that is particularly useful when the analyst wishes to analyze large population; it is obtained by smoothing a nonparametric estimator of the hazard rate defined on discrete time intervals.

This estimator has a variety of potential uses. First, it allows one to assess the qualitative shape of the hazard for a homogeneous population. The observed shape of the hazard can then be used to guide the selection of an appropriate parametric model. Second, it can be used in exploratory analyses to examine the effects of covariates on the hazard rate, for example, to informally assess whether a covariate affects the rate in a proportional way. Third, it can be used in graphical diagnostics performed after confirmatory analyses, for example, to assess departures from the assumed parametric form for a hazard specification that includes time or age dependence. Fourth, because the estimator can be easily graphed, it allows easy visual inspection of the hazard rate. This facilitates the discovery of unexpected aspects of the data and can help convey to the reader additional insight into the process under study.

This chapter begins with a brief overview of terms useful in event history analysis. I then discuss some common nonparametric estimators for the survivor, integrated hazard, and hazard functions in Section 2. Section 3 introduces a nonparametric estimator of the hazard and discusses issues related to smoothing nonparametric estimators of the hazard rate. Section 4 concludes the chapter by outlining a bootstrap procedure for obtaining confidence intervals for the smoothed hazard estimator of Section 3.

1.1 *Basic Terms*

For concreteness, I present statistical definitions in terms of the outcome of interest—a woman's age at first marriage. Let T_i° be an independent and continuous positive random variable representing the age at first marriage for individual i in a sample of I women, where $i = 1, \ldots, I$. We observe (or gather retrospective information) for individual i during the time interval $[0, \tau_i]$, where the τ_i represents the censoring time for individual i, for example, the age of the individual at the time of the last survey date for which information exists for individual i. Because not all individuals marry during the observation period, the observed data consist of a *final time* $T_i = \min(T_i^\circ, \tau_i)$, representing the age at first marriage or censoring, and a *censoring indicator*, $c_i = I(T_i^\circ > \tau_i)$, equal to one if the data for individual i are censored and zero otherwise.

1.1.2 *Survival, Integrated Hazard, and Hazard Rate*

Three important statistical concepts are the *survivor*, *integrated hazard*, and *hazard rate* functions. The survivor function, $G_i(t)$, has a simple and intuitive definition—it is the probability

that individual i has not married by age t

$$G_i(t) = \Pr[T_i \geq t] \qquad (3.1)$$

where T_i denotes the age of the individual at first marriage. The hazard rate is defined as the instantaneous probability of an event at age t given that no event has yet occurred; that is,

$$r_i(t) = \lim_{\epsilon \downarrow 0} \frac{\Pr(T_i \leq t + \epsilon \mid T_i \geq t)}{\epsilon} \qquad (3.2)$$

where the limit is taken for $\epsilon > 0$. The integrated hazard function is another useful concept and is defined by

$$H_i(t) = \int_0^t r_i(s)ds \qquad (3.3)$$

The three concepts are closely related in the following manner

$$G_i(t) = \exp[-H_i(t)] = \exp\left(-\int_0^t r_i(s)ds\right) \qquad (3.4)$$

There are a number of advantages of using the hazard rate to model event histories; see Tuma and Hannan (1984) and Cox and Oakes (1985) for detailed discussions of these advantages. For the purposes of this exposition, it is sufficient to note that Equation (3.4) implies that one can easily derive the integrated hazard and survivor functions from the hazard rate, which demonstrates the greater generality of modeling event history processes in terms of the hazard rate. A disadvantage of such a strategy is that the hazard rate is is unobservable due to censoring and because it is defined as an instantaneous probability. This differs from multiple regression, where the dependent outcome is easily observed.

One can, however, obtain estimators for the survivor probability, integrated hazard, *and* hazard rate under minimal parametric assumptions. Section 2.1 discusses standard nonparametric estimators for the the survivor probability and integrated hazard; Section 2.2 discusses a nonparametric method for estimating the hazard rate.

2.1 Nonparametric Estimation: Survivor and Integrated Hazard Functions

The Kaplan-Meier (1958) and Nelson-Aalen (Nelson, 1972; Aalen, 1978) estimators provide two extremely useful means by which to estimate the survivor and integrated hazard functions, respectively, from empirical data. Both these estimators possess excellent properties; in particular, they can be shown to be the maximum-likelihood estimator of the survivor function and integrated hazard function for *homogeneous* populations when minimal assumptions are made about the distribution of event and censoring times. Because of their "distribution-free properties", these estimators are termed *nonparametric*.

Let $R(t)$ denote the *risk set* of individuals at risk of marriage just prior to age t

$$R(t) = \{i : T_i \geq t\} \qquad (3.4)$$

In the case of first marriage, the risk set consists of all individuals who have not married by age t. Let $dN(t)$ denote the number of individuals who marry at age t. Then the Kaplan-Meier estimator for the survivor function $G(t)$ is defined by

$$\widehat{G}_{\text{K-M}}(t) = \prod_{R(t)} 1 - \frac{dN(t)}{R(t)}, \tag{3.5}$$

where the product is taken over the set of all individuals still at risk of marriage just prior to age t. The Nelson-Aalen estimator for the integrated hazard $H(t)$ closely resembles the definition of the Kaplan-Meier estimator and is given by

$$\widehat{H}_{\text{N-A}}(t) = \sum_{R(t)} \frac{dN(t)}{R(t)} \tag{3.6}$$

where again the sum is taken over the set of individuals at risk of marriage just prior to age t.

2.2 Maximum Likelihood Estimators for the Hazard Rate

While the method of maximum likelihood is helpful for obtaining estimators for the survivor and integrated hazard functions, it is less so for obtaining a estimator for the hazard rate. For example, Equations (3.5) and (3.6) suggest two ways of estimating $r(t)$

$$\hat{r}_{\text{K-M}}(t) = -\log[1 - dN(t)/R(t)]$$

and

$$\hat{r}_{\text{N-A}}(t) = dN(t)/R(t)$$

The two estimators are closely related as can be seen by the Taylor series

$$-\log[1 - dN(t)/R(t)] = [dN(t)/R(t)] + \frac{[dN(t)/R(t)]^2}{2} + \frac{[dN(t)/R(t)]^3}{3} + \dots$$

but differ in that $0 \le \hat{r}_{\text{N-A}}(t) \le 1$ while $0 \le \hat{r}_{\text{K-M}}(t) < \infty$.

A drawback is that both of these definitions yield estimators that consist of "spikes" of height $-\log[1 - dN(t)/R(t)]$ or $dN(t)/R(t)$ at the event times. The behavior of these estimates is thus similar to the nonparametric maximum likelihood estimator for an unknown density.

Several procedures for dealing with the "spiky" nature of these hazard estimators have been investigated including kernel approaches (see, e.g., Ramlau-Hansen, 1983a,b; Yandell, 1983; Tanner, 1984; Tanner and Wong, 1984), spline approaches (Anderson and Senthilselvan, 1980), and actuarial/incidence rate approaches (Finnäs, 1980; Borgan and Ramlau-Hansen, 1985). The kernel and spline approaches have attractive statistical properties. However these methods typically require very substantial computational resources when the hazard rate has nonconstant curvature or when sample sizes are very large, although certain spline approaches show considerable promise (see especially Silverman, 1984, 1985). The computational complexity of kernel or spline methods can be a serious drawback when studying many discrete outcomes in the social

sciences. For example, we can reasonably expect that the hazard rate for first marriage first rises and then declines with age and thus may exhibit nonconstant curvature (see, e.g., Coale and McNeil, 1972; Hoem, 1972; more generally, we may expect that the hazard rates of many social and demographic outcomes—fertility, divorce, job tenure, duration of education—have nonconstant curvature with age or duration. Similarly, social scientists often have the luxury of large samples, often running to several thousand cases or more, and it is precisely in these circumstances that the analyst might reasonably attempt to use a nonparametric estimator for the hazard rate. Fortunately, the actuarial/incidence rate approach requires far less computational effort and a variant on this approach forms the basis for the estimation procedure described in Section 3.

3.1 *A Piecewise Constant Actuarial/Incidence Hazard Estimator*

An alternative way of obtaining a hazard rate estimator is motivated from histogram estimates of a density. Recall that the histogram "bins" the data into discrete intervals and uses the empirical frequencies to construct an empirical density estimator. The bin widths are often taken as a fixed value, and optimal bin widths can be shown to be a function of the sample size (Scott, 1979; Freedman and Diaconis, 1981).

This suggests, heuristically, the use of an estimator that is closely related to the traditional actuarial hazard estimator (see, for example, Cox and Oakes, 1984, pp. 53–55). Consider time intervals of width Δt_j of the form $[t_{j-1}, t_j)$, where t_0 denotes the start of the process, $j = 1, \ldots, J$, and J denotes the number of bins; note that this definition allows for time intervals of varying width. Then consider the estimation of a piecewise constant hazard rate for each time interval

$$\hat{r}_{\text{A-I}}(t) = \rho_j, \quad t_{j-1} \leq t < t_j \tag{3.7}$$

subject to a piecewise constant censoring mechanism

$$\hat{r}_{\text{censor}}(t) = \lambda_j, \quad t_{j-1} \leq t < t_j \tag{3.8}$$

where the λ_j are considered "nuisance" parameters.

As Cox and Oakes note, these assumptions are not entirely plausible since one rarely believes that the underlying hazard rate and censoring mechanism are piecewise constant functions. On the other hand, these assumptions are widely used in demography and the social sciences (see, e.g., Hoem, 1976; Michael and Tuma, 1985) since they allow one to construct an estimator that follows the data in a manner that avoids strong parametric assumptions while circumventing the difficult problem of estimating a *continuous* hazard rate in a distribution-free setting.

Suppose we observe R_j individuals at risk just prior to the start of the interval $[t_{j-1}, t_j)$. Because the data are binned, or because we obtain data only on a time scale of finite accuracy, many individuals may be censored or experience the event during the interval. Let dN_j denote the number of individuals who experience the event and dC_j denote the number of individuals who are censored in the interval $[t_{j-1}, t_j)$. Then there are three distinct contributions to the joint

likelihood of ρ_j and λ_j in the interval $[t_{j-1}, t_j)$. One contribution is from the $R_j - dN_j - dC_j$ cases that survive and are uncensored through the interval $[t_{j-1}, t_j)$; the probability of this circumstance for these cases conditioned on survival up to $t_j - 1$ is

$$\exp[-\Delta t_j (\rho_j + \lambda_j)]$$

where $\Delta t_j = t_j - t_{j-1}$. A second contribution is from the dN_j cases that experience the event, and the corresponding conditional probability is

$$\int_0^{\Delta t_j} \rho_j e^{\rho_j s} e^{\lambda_j s} ds = \frac{\rho_j}{\rho_j + \lambda_j} \{1 - \exp[-\Delta t_j (\rho_j + \lambda_j)]\}$$

The third contribution is from the dC_j cases that are censored, with conditional probability

$$\int_0^{\Delta t_j} \lambda_j e^{\rho_j s} e^{\lambda_j s} ds = \frac{\lambda_j}{\rho_j + \lambda_j} \{1 - \exp[-\Delta t_j (\rho_j + \lambda_j)]\}$$

Then the log of the joint likelihood is given by

$$\begin{aligned} L_j(\rho_j, \lambda_j) = &- (R_j - dN_j - dC_j)\Delta t_j (\rho_j + \lambda_j) + dN_j \log[\rho_j/(\rho_j + \lambda_j)] \\ &+ dC_j \log[\lambda_j/(\rho_j + \lambda_j)] + (dN_j + dC_j) \log\{1 - \exp[-\Delta t_j (\rho_j + \lambda_j)]\} \end{aligned} \qquad (3.9)$$

Differentiating and solving the maximum-likelihood estimates for ρ_j and λ_j yields

$$\hat{\rho}_j = -\frac{dN_j}{\Delta t_j (dN_j + dC_j)} \log\left(1 - \frac{dN_j + dC_j}{R_j}\right) \qquad (3.10)$$

and

$$\hat{\lambda}_j = -\frac{dC_j}{\Delta t_j (dN_j + dC_j)} \log\left(1 - \frac{dN_j + dC_j}{R_j}\right) \qquad (3.11)$$

Hence the piecewise constant actuarial/actuarial estimator is given by

$$\hat{r}_{\text{A-I}}(t) = \hat{\rho}_j = -\frac{dN_j}{\Delta t_j (dN_j + dC_j)} \log\left(1 - \frac{dN_j + dC_j}{R_j}\right), \quad t_{j-1} \le t < t_j \qquad (3.12)$$

The actuarial/incidence estimator is closely related to the traditional life table estimator

$$\hat{r}_{\text{L-T}}(t) = \hat{\rho}_{j,\text{L-T}} = \frac{dN_j}{\Delta t_j [R_j - (dN_j + dC_j)/2]}$$

since for small $(dN_j + dC_j)/R_j$, the life table estimator can be shown to be a good approximation to the actuarial/incidence estimator by expanding the quantity $\log[1 - (dN_j + dC_j)/R_j]$ in a Taylor series. Similarly, if observations are obtained on a fine enough time scale so that each interval contains the final time for one individual only, then the estimator in Equation (3.12) reduces to a quantity familiar from the Kaplan-Meier estimator of the survivor function

$$\hat{r}_{\text{KM}}(t) = \hat{\rho}_{j,\text{KM}} = (\Delta t_j)^{-1} \log(1 - dN_j/R_j)$$

3.2 Smoothing on the Logarithm of the Hazard Estimator

A difficulty with the estimator in Equation (3.12) is that the resulting estimate is "rough". Figure 3.1 illustrates $\hat{r}_{A-1}(t)$ for data taken from the June 1980 Current Population Survey for a sample of 50139 white women, aged 12 to 80. The resulting plot suggests that the estimated rate of first marriage might consist of a smooth underlying rate of first marriage coupled with sampling variability or "noise". More formally, let $\tilde{r}(t)$ be a smooth estimator of the rate of first marriage; then we wish to obtain $\tilde{r}_{A-1}(t)$ from $\hat{r}_{A-1}(t)$ in the following manner

$$\hat{r}_{A-1}(t_j) = \tilde{r}_{A-1}(t_j) + g(\epsilon(t_j))$$

where $\epsilon(t_j)$ is a stochastic component with a symmetric distribution centered about zero, for example, $\epsilon(t_j) \sim \text{Gau}(0, \sigma_j^2)$, and g is some monotonic transformation. To determine the transformation g requires knowledge of the sampling distribution for the smoothed hazard estimator in finite samples. Because $\hat{r}_{A-1}(t)$ is a maximum likelihood estimator, it has an asymptotic limiting Gaussian distribution (see, e.g., Hoem, 1976), but the exact finite sampling distribution of $\hat{r}_{A-1}(t)$ appears difficult to derive for general censoring mechanisms.

Lacking exact finite sample results, we can instead seek an approximate normalizing transformation g. Cox and Oakes (1984; see also Miller, 1980) suggest that the small sample distribution for the well known maximum likelihood estimator for the exponential distribution can be approximated by χ^2_{2dN}, a chi-square distribution on $2dN$ degrees of freedom, where dN denotes the observed number of events; a small Monte Carlo study reported by Lawless (1982) indicates that this approximation is reasonably accurate for $dN \geq 10$. This suggests, on purely heuristic grounds and by reasoning from the similarity of $\hat{r}_{A-1}(t)$ to the MLE for the exponential distribution, that the finite sample distribution of $\hat{\rho}_j$ can be roughly approximated by $\chi^2_{2dN_j}$.

The χ^2 distribution is highly skewed, particularly for small degrees of freedom. This suggests that smoothing directly on $\hat{r}_{A-1}(t)$ will introduce considerable upward bias for the smoothed estimate $\tilde{r}_{A-1}(t)$ when the number of events in a time interval is small. But various normalizing transformations for χ^2 variates exist, for example, the well-known Wilson-Hilferty transformation (see, e.g., Johnson and Kotz, 1970)

$$\chi^2_\nu(\alpha) = \nu\left(z_\alpha\sqrt{\frac{2}{9\nu}} + 1 - \frac{2}{9\nu}\right)^3$$

where $\chi^2_\nu(\alpha)$ denotes the α percentage point of the χ^2 distribution with ν degrees of freedom and z_α is the α percentage point of the standard Gaussian distribution.

Using the Cox-Oakes approximate distribution for $\hat{\rho}_j$ we can form the following $1 - \alpha$ confidence interval

$$\chi^2_{2dN_j}(1 - \alpha/2) < 2dN_j\rho_j/\hat{\rho}_j < \chi^2_{2dN_j}(\alpha/2)$$

where $\chi^2_{2dN_j}(\alpha)$ denotes the α percentage point of the χ^2 distribution with $2dN_j$ degrees of freedom. Utilizing the Wilson-Hilferty transformation and taking logs yields

$$3\log\left(1 - \frac{1}{9dN_j} - \frac{z_{\alpha/2}}{3\sqrt{dN_j}}\right) < \log\rho_j - \log\rho_j < 3\log\left(1 - \frac{1}{9dN_j} + \frac{z_{\alpha/2}}{3\sqrt{dN_j}}\right)$$
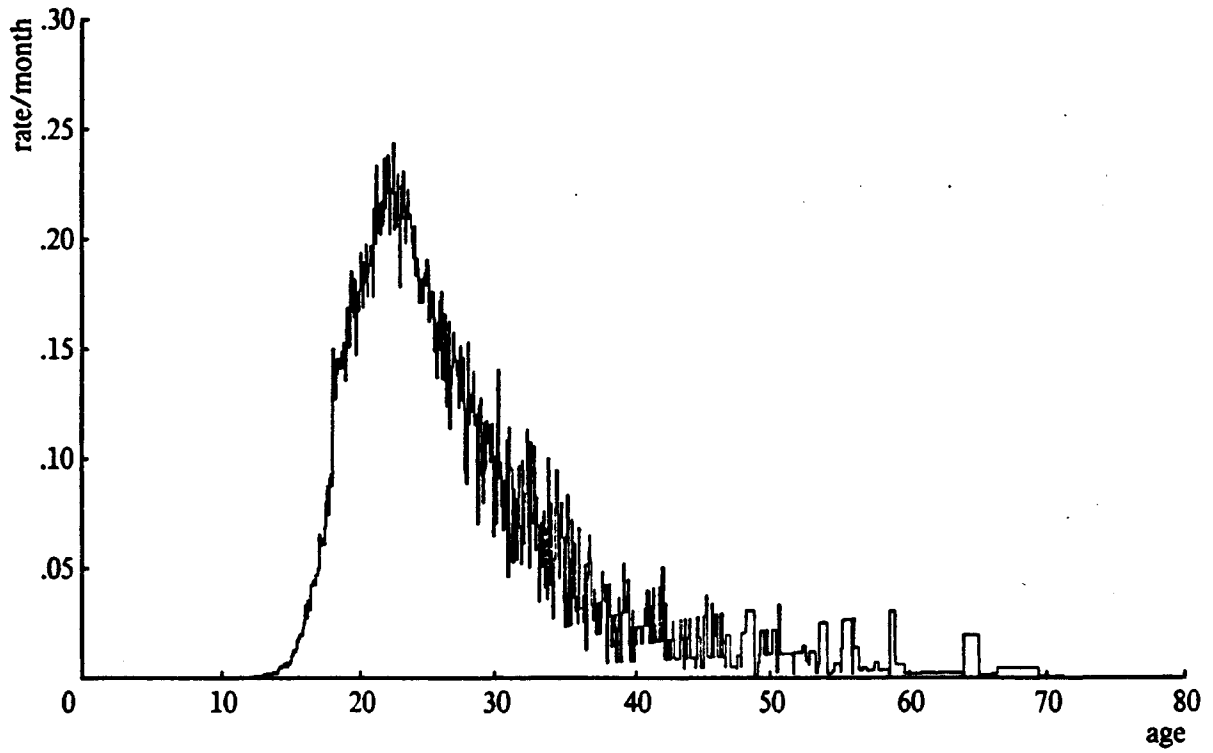
Figure 3.1: Actuarial/incidence Hazard Estimates $\hat{r}(t)$: Whites, $I = 50139$ (June 1980 CPS). Time intervals $[t_{j-1}, t_j)$ are of width $\Delta t = t_j - t_{j-1} = $ one month.
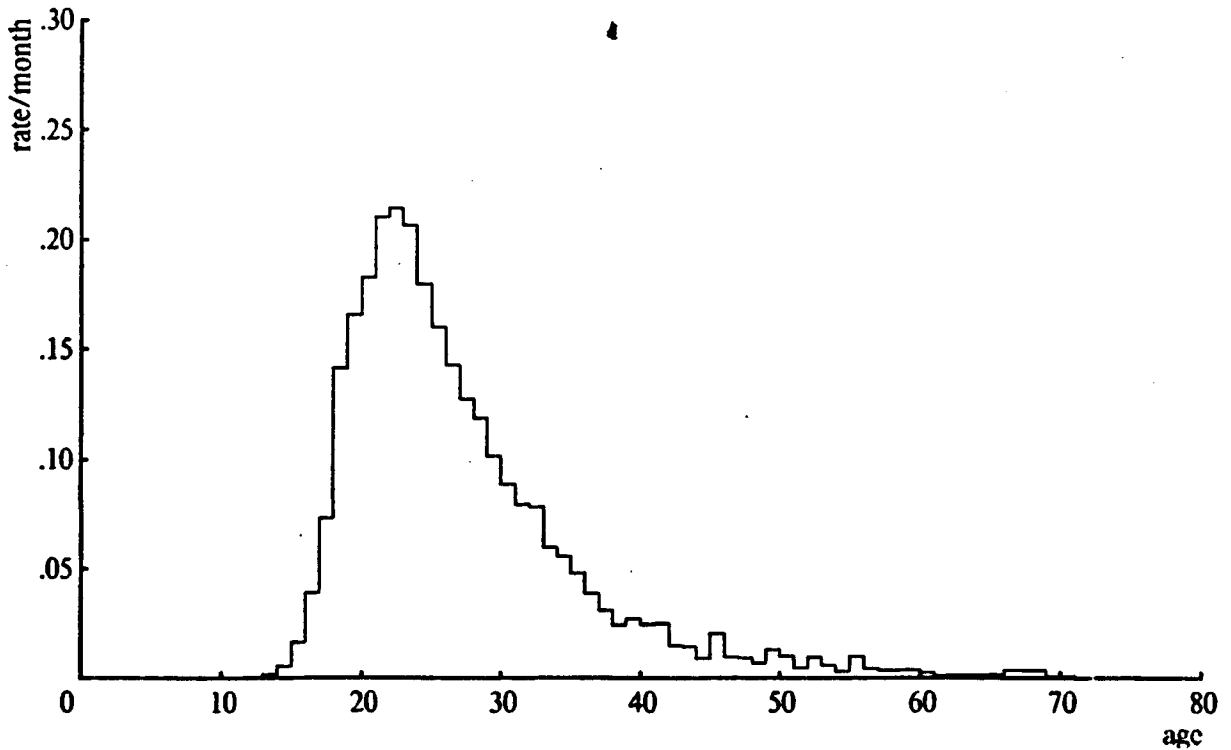
Figure 3.2: Actuarial/incidence Hazard Estimates $\hat{r}(t)$: Whites, $I = 50139$ (June 1980 CPS). Time intervals $[t_{j-1}, t_j)$ are of width $\Delta t = t_j - t_{j-1} = $ one year.

Then expanding the lower and upper percentiles in a Taylor series and neglecting all terms of second order and higher yields

$$\frac{-z_\alpha}{\sqrt{dN_j}} < \log \rho j - \log \hat{\rho}_j + \frac{1}{3dN_j} < \frac{z_\alpha}{\sqrt{dN_j}}$$

These heuristics suggest that, to a first order approximation,

$$\sqrt{dN_j}(\log \rho_j - \log \hat{\rho}_j) \rightarrow \text{Gau}(0, \sigma_j^2) + O(1/\sqrt{dN_j})$$

and hence that one may smooth more reliably on the logarithm of the actuarial/incidence rate than on the raw actuarial/incidence rate. This provides an informal justification for an assertion by Miller (1981) that the limiting distribution for the logarithm of a constant hazard rate estimator is more nearly Gaussian than that for the untransformed estimator.

3.3 Smoothing Procedures

There are several possible smoothing procedures that one might employ. One possibility is to choose a larger value for Δt_j. This results in a somewhat less jagged plot, but at the cost of "bumpiness." See Figure 3.2, which replicates Figure 1 with Δt_j equal to one year. A second procedure is to consider flexible parametric specifications; see, for example, Hoem (1972) and Hoem et al. (1980).

Another method for smoothing is to use a local average of the $\hat{r}_{A-1}(t)$'s in a neighborhood of the time interval t_j as the smoothed value for the hazard rate. The size of the neighborhood, or span, over which the averaging is taking place controls the degree of smoothing. Larger spans, corresponding to larger neighborhoods, result in a smoother overall estimate but increases bias; smaller spans result in less bias but increases the variability of estimates. A convenient choice of the span is to take a symmetric neighborhood consisting of a total of $2K + 1$ points, K to the left and right of the point t_j. Typically, the number of points, $2K + 1$, in the neighborhood ranges between 5 to 50 percent of the data. It is not possible to maintain symmetric neighborhoods near the endpoints; one solution (Friedman and Tibshirani, 1984) is to shrink the number of points in an asymmetric way, for example, by taking the indices k for the points t_k of the span neighborhood of t_j by $k = \max(1, j - K), \ldots, \min(j + K, K)$, where $j = 1, \ldots, J$ and K denotes the span parameter.

While the local averaging smoother is attractive in its simplicity, it has certain drawbacks. In particular, if the t_j's are not equally spaced, the local averaging smoother is unable to reproduce a linear relationship for pairs of points $(t_j, \hat{r}(t_j))$ that lie on a straight line (Friedman, 1984). At first glance this may not seem to pose a problem for the estimate of $\hat{r}(t_j)$ proposed in Equation (3.12). However, it is likely that the observed number of marriages will fluctuate more greatly with increasing age due to sampling variability and the lower probability of marriage at older ages. Because of this, it is advantageous to define time intervals with a larger Δt_j at later ages to deal with the sparseness of events, with the result that the t_j may be unequally

spaced at later ages. A second drawback of the local averaging smoother is that it is subject to bias near the endpoints. If $\hat{r}_{\text{A-1}}(t)$ is positively (negatively) sloped near an endpoint, the local averaging smoother will produce estimates that are downwardly (upwardly) biased (Cleveland, 1979; Friedman, 1984).

Cleveland (1979) and Friedman (1984) note that these two drawbacks can be overcome by fitting a local least squares line to the values of $\hat{r}(t_j)$, i.e., by estimating the simple x–y regression of the form $\hat{r}_{\text{A-1}}(t) = \hat{\rho}_j \alpha_{j,K}(t_k) + t_j \beta_{j,K}(t_k)$, where the subscripts j and K for the local regression parameters $\alpha_{j,K}(t_k)$ and $\beta_{j,K}(t_k)$ denote that the regression is taken for the K points t_k to the left and right of t_j. Then in the spirit of the local averaging smoother, one can define the *local linear smoother* by using the predicted value $\tilde{r}_{\text{A-1}}(t) = \tilde{\rho}_j = \hat{\alpha}_{j,K}(t_k) + t_j \hat{\beta}_{j,K}(t_k)$ obtained by fitting the least squares line to the K points to the left and right of the point t_j for a total of $2K + 1$ points in the regression, where $\tilde{r}$ and $\tilde{\rho}$ denote smoothed estimates as opposed to the original estimates $\hat{r}$ and $\hat{\rho}$. The local linear smoother produces estimates that are identical to the local averaging smoother for equally spaced data away from the endpoints, but with less bias near the endpoints. It can also reproduce linear relationships for unequally spaced data. The use of updating formulas makes the computational complexity of the local least squares smoother nearly the same as local averaging smoother.

3.3 A Variable Span Local Linear Smoother

As noted above, the choice of the span represents a trade-off between bias and variability of the estimate. In addition, a fixed span value is not optimal if the underlying hazard rate has a nonconstant curvature (a nonconstant second derivative) or if the variability of estimates increases or decreases with age. If either of these situations occurs, then it is desirable to allow the span value to vary.

Friedman (1984) proposes a variable span smoother based on linear least squares fits in which the span is chosen in an adaptive manner according to local features of the data. The estimator can be described in terms of a five part algorithm. First, the estimator computes three primary smooths using three fixed spans with $2K + 1$ chosen as $0.05J$, $0.2J$, and $0.5J$, respectively. These spans thus capture high, middle, and low frequency information in the data. Second, the algorithm computes cross-validated absolute residuals for each fixed span and smooths the each set of residuals against the the primary smooths using the $0.2J$ span. Third, the span minimizing the sum of cross-validated and smoothed residuals in step two is used as a preliminary "best" smooth for each t_j. This step also contains provisions for biasing the span towards larger values so as to produce a smoother final estimate. Fourth, the preliminary span values from step three are themselves smoothed using the $0.2J$ span in order to limit the variability of estimated optimal spans. Last, the smoothed span values are used to construct the final smoothed predicted value by interpolating between the primary smoothed values.

Because rates of first marriage are likely to exhibit both curvature (changing second derivatives) and nonconstant variability with age, I have chosen to use Friedman's estimator to obtain

smoothed estimates for rates of first marriage. While the estimator is clearly more complicated than the local averaging or local linear smoothers, the algorithm remains linear in the number of binned data points, J, and so is extremely fast in practice, in part because the number of binned data points is far less than the number of individuals in the data. For example, in estimating the hazard rate in Figure 3.1, the number of time intervals defined by age in months is less than 500 for a sample of over 50000 white females in the U.S. for data taken from the June 1980 Current Population Survey.

4.1 Bootstrap Confidence Intervals for Smoothed Hazard Estimates

As a matter of principle, one would like to accompany an estimate of a hazard rate by some assessment of the variability of this estimate. This task is particularly important when attempting to make comparisons between groups. In a parametric setting, one would typically conduct analyses using an estimator whose properties are well understood for some assumed parametric sampling distribution. Then the setting of confidence intervals for an estimate can be accomplished by using standard parametric methods. But the task of assessing the variability of an estimate is greatly complicated in analyses in which we do not know, or are unwilling to make assumptions about, the exact parametric sampling distribution for the data. And because of the nonparametric nature of the estimator presented in Section 3 and the smoothing procedure employed, assessing the variability of the smoothed hazard estimator by an analytic approach presents a difficulty task.

The bootstrap (Efron, 1979, 1982) provides a procedure that allows one to abandon many usual parametric assumptions while obtaining estimates for the variability of an estimator that have excellent statistical properties. Specifically, the bootstrap provides a highly general nonparametric procedure for assessing the quality (variability and bias) of an estimate by "reusing" the observed data. It is both more general and more efficient than other nonparametric techniques and represents a departure from more traditional parametric procedures by substituting the computing power for the parametric sampling assumptions of classical methods. This section outlines a bootstrap procedure used to set pointwise confidence intervals and assess bias for the smoothed hazard estimator described in Section 3.

4.2 Bootstrapping Censored Data

As in Section 1.1, let T_i and c_i be the observed final time and censoring indicator, respectively, for individual i in a sample of I individuals. In keeping with the assumptions underlying the piecewise actuarial/incidence estimator, we assume that the corresponding event and censoring times T_i° and τ_i can be characterized by piecewise constant hazard rates ρ_j and λ_j in the time interval $[t_{j-1}, t_j)$.

More formally, suppose $T_1^\circ, \ldots, T_I^\circ$ are iid event times drawn from an unknown survivor distribution $G(t)$ and suppose $\tau_1, \ldots, \tau_I$ are iid censoring times drawn from an unknown censoring survivor distribution $H(\tau)$. We observe $T_i = \min(T_i^\circ, \tau_i)$ and a censoring indicator $c_i = 1$ if the data for individual i are censored and 0 otherwise. Let $\tilde{\rho}_j(T_1, \ldots, T_I)$ and $\tilde{\lambda}_j(T_1, \ldots, T_I)$ denote

the (functional) estimators defined in Equation (3.11) for the unknown parameters ρ_j and λ_j for a sample of I individuals and let $\widehat{G}(t)$ and $\widehat{H}(\tau)$ be the sample survivor distributions for the event and censoring times constructed by applying the Kaplan-Meier estimators to the observed event and censoring times, respectively, for this sample.

We wish to set confidence intervals for the estimates $\tilde{\rho}_j$ obtained by applying the variable span smoother to the $\hat{\rho}_j$. To do so using standard techniques requires knowledge of the sampling distribution of $\tilde{\rho}_j$ for the unknown survivor distributions G and H. This task is complicated by the nature of the smoothing algorithm and the quality of the log transformation to approximate normality for varying numbers of events dN_j in each time interval. The bootstrap method approximates the sampling distribution of $\tilde{\rho}_j$ in a nonparametric way by estimating the sampling distribution of $\tilde{\rho}_j$ for the *observed* distributions $\widehat{G}$ and $\widehat{H}$, which lends itself to a Monte Carlo procedure such as the following:

1. Construct $\widehat{G}$ and $\widehat{H}$ from the observed data.

2. Let p_e denote the observed proportion of individuals with observed events. Then construct a *bootstrap* sample of observed final times, $T_1^*, \ldots, T_I^*$, by flipping a biased coin with Pr(heads)= p_e. If the coin toss yields a heads, then sample the final time, T_i^*, with replacement from $\widehat{G}$ and put $c_i^* = 0$; otherwise sample T_i^* with replacement from $\widehat{H}$ and put $c_i^* = 1$. The superscript $*$ emphasizes that the realized values are bootstrap values sampled with replacement from the observed distributions.

3. Iterate on step (2) B times. This yields estimates $\hat{\rho}_{j,b}^*(T_1^*, \ldots, T_I^*)$, where $b = 1, \ldots, B$, and $j = 1, \ldots, J_b$. The notation J_b emphasizes that the number of bins may vary across bootstrap iterations if, for example, the time intervals are chosen according to some data dependent rule.

4. Form the bootstrap sampling distribution for $\tilde{\rho}_j$, $\widehat{CDF}_j(y) = \#(\hat{\rho}_{j,b}^* \le y)/B$. One can then use $\widehat{CDF}_j(y)$ to form approximate bootstrap confidence intervals for $\tilde{\rho}_j$.

This Monte Carlo procedure can be seen to be equivalent to sampling, with replacement, from the observed pairs of final times and censoring indicators (T_i, c_i). Thus one can employ the following equivalent bootstrap procedure

1. Construct bootstrap data of the form $(T_{i,b}^*, c_{i,b}^*)$, $i = 1, \ldots, I$, by sampling with replacement from the observed data (T_i, c_i). Calculate the bootstrap estimate $\tilde{\rho}_{j,b}^*$, where $j = 1, \ldots, J_b$.

2. Repeat step (1) B times, which yields estimates $\tilde{\rho}_{j,b}^*$, where $b = 1, \ldots, B$.

3. Form the bootstrap sampling distributions for the $\tilde{\rho}_j$, $\widehat{CDF}_j(y) = \#(\tilde{\rho}_{j,b}^* \le y)/B$. Approximate bootstrap confidence intervals by inverting $\widehat{CDF}_j$.

4.3 A Simple Bias Correction

The technique of using of log transformation to correct the skewness in the distribution of the $\hat{\rho}_j$'s can be expected to work reasonably well when there are a sufficient number numbers of events dN_j in the time interval $[t_{j-1}, t_j)$. Then the variable span smoother can be used with confidence to obtain smooth estimates $\tilde{\rho}_j$. However, the various approximations used in the

informal derivation of this result, particularly the Cox-Oakes χ^2 approximation for the sampling distribution of $\hat{\rho}_j$, is likely to be less good when the observed number of events in a time interval is small. Because the smoother relies on a reasonably symmetric distribution of the ρ_j's to produce a smooth curve, the smoothed estimates of $\tilde{\rho}_j$ are subject to potential bias, particularly at later ages when there are fewer observed marriages. This section considers methods to estimate bias and produce estimates correcting for bias when obtaining $\tilde{\rho}_j$ and confidence intervals for $\tilde{\rho}_j$.

Consider the following measure of bias for $\tilde{\rho}_j$

$$\text{Bias}(\tilde{\rho}_j) = E_{G,H}[\tilde{\rho}_j(T_1, \ldots, T_I)] - \rho_j \tag{3.13}$$

where G and H are the unknown survivor distributions, $E_{G,H}[\tilde{\rho}(T_1, \ldots, T_I)]$ denotes the expectation of the estimate $\tilde{\rho}_j$ for the unknown G and H, and ρ_j denotes the true (smooth) population value. Efron (1981) suggests the following bootstrap estimator for the quantity in Equation (3.13)

$$\text{Bias}^*(\tilde{\rho}_j) = E^*_{\widehat{G},\widehat{H}}[\tilde{\rho}_j(T_1^*, \ldots, T_I^*)] - \tilde{\rho}_j(T_1, \ldots, T_I)$$

or, equivalently,

$$\text{Bias}^*(\tilde{\rho}_j) = \frac{1}{B} \sum_{b=1}^{B} \tilde{\rho}_{j,b}^* - \tilde{\rho}_j \tag{3.14}$$

where the expectation is taken over bootstrap samples obtained from the observed distributions $\widehat{G}$ and $\widehat{H}$. A closer inspection of $\text{Bias}(\tilde{\rho}_j)$ and $\text{Bias}^*(\tilde{\rho}_j)$ shows that Bias^* differs from Bias by substituting $\widehat{G}$ and $\widehat{H}$ for the unknown G and H. Clearly, this estimator cannot be defended on general grounds, but is, roughly speaking, a reasonable estimator of bias when the survivor distributions G and H are "close" to the observed distributions $\widehat{G}$ and $\widehat{H}$. See Schenker (1985) for an example in which the estimator of bias breaks down and Parr (1985) for a more comprehensive discussion of related robustness issues in bootstrapping.

Given an estimator of bias, one can then form simple bias corrected confidence intervals in the following manner (Efron, 1986). For a coverage probability α, let

$$\tilde{\rho}_j^{\text{Low}}(\alpha) = \tilde{\rho}_j - \text{Bias}^*(\tilde{\rho}_j) + [\widehat{CDF}^{-1}(1 - \alpha/2) - \widehat{CDF}^{-1}(1/2)] \tag{3.15}$$

and

$$\tilde{\rho}_j^{\text{High}}(\alpha) = \tilde{\rho}_j - \text{Bias}^*(\tilde{\rho}_j) + [\widehat{CDF}^{-1}(\alpha/2) - \widehat{CDF}^{-1}(1/2)] \tag{3.16}$$

Then an approximate $1 - \alpha$ confidence interval for $\tilde{\rho}_j$ is given by

$$[\tilde{\rho}_j^{\text{Low}}(\alpha), \tilde{\rho}_j^{\text{High}}(\alpha)]$$

An Exploratory Look at Age Dependence in Rates of First Marriage

This chapter presents exploratory analyses of first marriage patterns for a sample of women taken from the June 1980 Current Population Survey. These results provide one of the first serious and detailed empirical examinations of the pattern of age dependence in hazard rates of first marriage by utilizing nonparametric statistical methods that allow the direct visual assessment of how hazard rates of marriage depend on age and measured individual attributes.

The goals of this chapter are simultaneously modest and ambitious. They are modest inasmuch as the results presented are exploratory and suggestive rather than theory-driven and confirmatory. For example, in the analyses of this chapter I examine the effects of only a few covariates on the rate of first marriage and hence ignore many covariates that are commonly hypothesized to affect the decision to marry. Moreover, the exploratory results are univariate and bivariate in nature—that is, I examine the effects of at most two covariates on the rate of first marriage in any given analysis—and thus overlook many possible sources of observable heterogeneity. Because of these limitations, the results reported in this chapter cannot be interpreted as definitive.

Far more ambitiously, I seek to answer the following questions, which I believe have a certain degree of theoretical importance. First, can one find empirical evidence that the pattern of age dependence in rates of marriage exhibits strong regularities, and what statistical techniques might prove useful in such a search? Second, if strong regularities are found, can they in turn be described in simple qualitative terms? Third, if observable regularities exist in the pattern of age dependence in rates of first marriage, do they in turn vary with individual characteristics that are thought to influence the social processes underlying marriage? Fourth, if strong regularities can be found, what informal assessments concerning the relative magnitude, direction, and significance of effects for different covariates can be supported by the data? Last and assuming points 1–4, what might the observed empirical regularities suggest about hypotheses for age patterns in rates of first marriage derived from different theoretical perspectives, and, more generally, what theoretical implications can be gleaned from the empirical evidence.

The chapter begins with a description of the data analyzed in this chapter. I then give an informal overview in Section 4.2 of the statistical techniques that were first introduced in Chapter 3 and which are used extensively throughout this chapter. Sections 4.3 and 4.4 present the empirical results and the substantive discussions of these results, respectively. Section 4.3, which comprises the majority of pages in this chapter, is itself divided into three sections. Section 4.3.1 presents a several nonparametric "looks" at race-ethnic differences for whites, blacks, and women of Mexican descent via estimates for the survivor, integrated hazard, hazard, and log hazard functions for first marriage. Sections 4.3.2 and 4.3.3 proceed at a more rapid pace and present exploratory results for the pattern of age dependence in log rates of first marriage for four broadly defined birth cohorts and four groups defined by level of completed education.

I have deliberately postponed all substantive discussion of the empirical results until Section

4.4, which concludes the chapter, and this choice merits special comment. A more usual orga-
nization would consist of a theoretical section setting out substantive hypotheses derived from
existing theory, followed by a set of confirmatory analyses, with a concluding section discussing
the results and providing speculations about further directions for research. I have chosen not
to follow such a standard format for three reasons. First, given our relative lack of knowledge
about the *specific* nature of hazard rates of first marriage and a corresponding lack of theo-
retical literature that speaks directly to rates of first marriage, I believed that the intensive
and detailed empirical investigations of Section 4.3 could better communicate to the reader a
sense of the ways in which the (log) hazard rate of first marriage varies with individual social
characteristics than would a theoretical section containing bald assertions about the effects of
individual attributes on an individual's rate of first marriage, as "derived" from existing theory.
Second, the empirical results in Section 4.3 were naturally organized around particular covariates
(race-ethnicity, birth cohort, and education), but the few substantive hypotheses that could be
inferred from the existing literature on marriage are best evaluated by a consideration of results
spanning the subsections of Section 4.3. Third, postponing a substantive discussion permits
a complete overview of the results of Section 4.3, and this in turn helps shed light on where
existing theory is and is not informative. Lastly, the nonstandard format reports in a faithful
manner the way in which I conducted the analyses and thus helps to emphasize the exploratory
and inductive nature of the results and discussions presented in this chapter.

4.1.1 *Data*
The data analyzed in this chapter are well suited to both the modest and ambitious goals
outlined in the introduction to this chapter. A drawback of these data are the lack of detailed
covariate information for individual respondents, and the variables I examine are limited to race-
ethnicity, birth cohort, and completed education. But despite limited information on individual
characteristics, these data provide sufficient retrospective information from which to construct
individual marriage histories for an extremely large sample of women. These aspects of the data
greatly facilitate the use of nonparametric event history methods that allow the examination of
age dependence in rates of first marriage without making a priori assumptions about the nature
of age dependence.

4.1.2 *Sample Design*
The Current Population Survey (CPS) is a monthly survey conducted by the United States
Bureau of the Census and is intended to provide estimates for labor statistics.[1] The sample in
the June Supplement to the 1980 Current Population Survey consists of a national probability
sample of approximately 77,000 households drawn from the civilian noninstitutional population
of the United States aged 14 or older in all 50 states and the District of Columbia. Households
are surveyed for a total of eight months according to a rotation schedule. The data for each

---

[1] The information in this section is taken from Carr, Steel, and Card, 1983. See also *The Current Population Survey: Design and Methodology*, Technical Paper No. 40, U.S. Bureau of the Census, 1977.

monthly supplement to the CPS are collected during the week containing the 19th day of the month.

Each monthly supplement to the CPS includes a basic set of questions pertaining to the individual's labor force activity, household composition, and information concerning the personal characteristics of each member of the household (relationship to the head of the household, date of birth, age, marital status, race and ethnic background, sex, level of education). The June Supplements for 1975 and 1980 include additional questions pertaining to the marital and fertility histories for all survey respondents. Marital histories are gathered for all male and female respondents aged 15 and older who have ever married; fertility histories are gathered for all individuals aged 18 and older and for individuals aged 15 and older if they have ever married.[2] Because of the wide range of ages covered by the June Supplement to the 1980 CPS, retrospective marital histories are available for women born between the years of 1880 to 1965.

4.1.3 Retrospective Event History Information

For the analyses conducted in this chapter, I utilized a subsample of 71,407 women drawn from the 77,000 household records of the June 1980 CPS.[3] These data include retrospective information from which to construct first marital histories, including data on a woman's date of birth (in month and year), age in years, marital status at the time of the June 1980 interview (never married, currently married spouse present, currently married spouse absent, separated, divorced, widowed), number of marriages, and date (in month and year) of first, second, and last marriage. I performed several checks on the consistency of recorded respondent responses including cross checks on a woman's number of marriages, marital status, current marital status and dates of marriages; and for a woman's age in years, date of birth, and age at first marriage. But because the data are gathered in one interview only, many discrepancies were impossible to reconcile and only limited imputation of data was possible. After reconciling and imputing where possible, approximately 15% of 71,407 cases were rejected because of data inconsistencies, leaving a sample of 60,737 women for whom first marriage histories could be reliably determined.

I was successful in determining, reconciling, and, where necessary, imputing the year and month of birth and first marriage for a large proportion of cases in the sample from information on a woman's age in years, date of birth, the date of first marriage, marital status, and number of marriages. I chose 12 years of age as an truncation point for age at marriage and rejected all cases with a reported first marriage age less than 12; this arbitrary truncation point affected only a small proportion of the sample and, indeed, very few women have recorded first marriage ages less than 15. There were a large number of discrepancies between reported age and reported date of birth, although most discrepancies amounted to a one year difference between the reported age at interview and that calculated from the reported year and month of birth. In most of these

---

[2] Thus fertility histories were not gathered for women less than 18 years old if they had never married.

[3] These raw data extracts (U.S. Bureau of the Census, 1983) were kindly made available to me by the Data Archive on Adolescent Pregnancy and Pregnancy Prevention.

cases, I used the reported month and year of birth. Some cases were missing information for month of birth but had a valid year of birth. I imputed a month of birth for these individuals by using the month of birth for the previous individual in the data file (or the last valid month of birth processed in the data file); this procedure mimics, in a rough manner, a random imputation procedure based on the observed frequency for month of birth in the sample. A somewhat larger proportion of cases were missing both month and year of birth but had a valid age at interview. I imputed the year of birth from age at interview and imputed the month of birth using the month imputation procedure described above. Data for the month and year of first, second, and last marriage were occasionally recorded in a manner that appeared to violate the questionnaire skip patterns; I reconciled these data whenever possible. Assigning censoring times was for the most part straightforward, and women who indicated a never married status, no marriages, and who had no valid dates of marriage were censored at the date of interview and assigned censoring ages equal to their age at interview.

I found it considerably more difficult to reconcile data relevant to a woman's marital status for these data; approximately 10.5% of the sample had no valid date of marriage and no valid number of marriages but reported some married status, or had some valid date of marriage or a valid number of marriages but reported a never married status. I subjected the latter cases to individual inspection but could usually do little more than impute a special "ever" marital status for these cases. Cases reporting some married status but missing a valid date of marriage or number of marriages presented similar difficulties because I could neither reliably impute a marriage date nor a censoring time. These cases constituted the majority that were rejected in constructing a "clean" first marital history file.

4.1.4 Available Covariate Information

As noted above, although the June CPS provides sufficiently detailed individual-level data to reconstruct retrospective event histories for a very large sample of women, there is only limited covariate information available for each respondent. Table 4.1 lists the variables that I have selected to examine along with information used to construct the first marital histories.

From the variables listed in Table 4.1 I constructed measures for race-ethnicity, birth cohort, and completed education that were used in all subsequent exploratory analyses. Because information on race and ethnicity were gathered in separate questions, I constructed a race-ethnicity variable by collapsing the categories of Mexican-American, Chicano, and Mexican into a single category for women of Mexican descent and collapsing the categories of for South and Central America into a single category. I then coded a woman's race-ethnicity based first on the ethnicity questionnaire item and second on the race questionnaire item if her recorded response to the ethnicity item was "other". Thus a person was coded as white or black only if her recorded response to the ethnic questionnaire item was "other" and her recorded response to the race questionnaire item was white or black. This coding procedure yielded 8 categories: white, black, Mexican descent, Puerto Rican, Cuban, Central/South American, other Spanish,

Table 4.1 June Supplement to the 1980 Current Population Survey: Selected Variables

Variable	Description
date of birth	day, month, year
age	in years
current marital status	married civilian spouse present
	married military spouse present
	married spouse absent
	widowed, divorced
	never married
number of times married	
date of marriage	month, year of first, second, last
race	white, black, other
ethnicity	Mexican American, Chicano, Mexican,
	Puerto Rican, Cuban, Central or South
	American, Other Spanish, Other ethnicity
highest grade completed	grades 0, 1, ..., 20+

and other ethnicity.

I used data on year of birth to create four large cohorts of women born between 1880–1909, 1910–1929, 1930–1949, and 1950–1965. Respondents were asked about the highest grade attended and if she had completed this grade; note that both items are current for the date of interview and so have a somewhat ambiguous interpretation in the context of first marriage especially for women who continued their education after marriage. I coded a completed education variable by calculating the highest grade completed and then collapsing responses into grades 0–9, 10-12, 13, and 14 or more. The separate category of 13 years of education (one year of college) requires some explanation. A preliminary examination of frequencies revealed that a large number of women reported that their highest educational level completed was one year of college. Because I expected that these women might differ from those who completed two or more years of college, I coded a separate category for these women.

4.2 Exploratory Methods for Event History Analysis

This section presents a brief and self-contained description of the smoothed hazard estimator used extensively throughout this chapter. Readers who desire a more detailed and leisurely description of the methods used should see Chapter 3. Other readers who are primarily interested in the exploratory results and substantive implications of these results may wish to proceed directly to Sections 4.3 and 4.4, taking the estimator and confidence intervals on faith.

Although there are a number of standard theoretical perspectives that inform research on marriage, existing theories of marriage offer little or no guidance about the specific shape of the rate of first marriage. In such a situation, nonparametric methods have practical attrac-

tions because they make minimal assumptions about the underlying distribution of event times and hence about the form of age dependence in rates of first marriage. Chapter 3 introduced a nonparametric estimator for the hazard rate obtained by smoothing an actuarial/incidence estimator for the hazard rate, where the hazard is assumed to take a constant value on discrete time intervals. The resulting smoothed hazard estimator can be plotted and used to make comparisons between groups differing on observed characteristics. An important feature of this estimator is that it allows the analyst to visually assess how patterns of age dependence in rates of first marriage differ with individual attributes.

I present the discussion in terms of the outcome of interest—the hazard rate of first marriage for women, with a particular emphasis on age dependence in the hazard rate. Consider pre-specified discrete time intervals of the form $[t_{j-1}, t_j)$ for J intervals; by convention $t_0 = 0$, where "time" is a woman's age. Then consider estimating a constant hazard ρ_j subject to a constant censoring mechanism λ_j in the time interval $[t_{j-1}, t_j)$ for each of the J time intervals, where the λ_j are considered "nuisance" parameters. This approach has the advantage of allowing the ρ_j to closely follow the data in a manner that makes minimal parametric assumptions about the global shape of the hazard rate $r(t)$.

The estimation problem described above lends itself to maximum likelihood techniques and one can write the joint likelihood of ρ_j and λ_j (see, e.g., Cox and Oakes, 1984), which yields the following maximum likelihood estimator for ρ_j

$$\hat{\rho}_j = -\frac{dN_j}{\Delta t_j (dN_j + dC_j)} \log\left(1 - \frac{dN_j + dC_j}{R_j}\right) \tag{3.12}$$

where $\Delta t_j = t_j - t_{j-1}$, dN_j denotes the number of women who marry in the interval $[t_{j-1}, t_j)$, dC_j the number of women who are censored in the interval, and R_j the number of women at risk of marriage just prior to the start of the interval.

For small Δt_j, the estimates $\hat{\rho}_j$ exhibit substantial stochastic fluctuation, which suggests that the $\hat{\rho}_j$ can be represented by a smooth component $\tilde{\rho}_j$ plus a random "noise" component

$$\hat{\rho}_j = \tilde{\rho}_j + g(\epsilon_j)$$

where the ϵ_j are assumed to have some symmetric distribution and g is some monotonic transformation. Some heuristic considerations in Chapter 3 suggest that the sampling distribution of the stochastic component is more nearly Gaussian for the logarithm of the $\hat{\rho}_j$, which suggests that standard smoothing techniques can be applied to the log $\hat{\rho}_j$ to obtain the smooth estimates $\tilde{\rho}_j$

$$\log \hat{\rho}_j = \log \tilde{\rho}_j + \epsilon_j$$

where the ϵ_j are approximately distributed according to $\text{Gau}(0, \sigma_j^2)$.[4]

---

[4] As noted in Chapter 3, the approximate normality of the ϵ breaks down when the number of events dN_j in the interval $[t_{j-1}, t_j)$ is very small. Because of this, I have chosen the intervals $[t_{j-1}, t_j)$ in the analyses of Section 4.3 such that each interval has at least observed ten events.

A number of techniques have been recently suggested for obtaining smooth hazard estimates including kernel approaches (see, e.g., Ramlau-Hansen, 1983a,b; Yandell, 1983; Tanner, 1984; Tanner and Wong, 1984), spline approaches (Anderson and Senthilselvan, 1980), and actuarial/incidence rate approaches (Finnäs, 1980; Borgan and Ramlau-Hansen, 1985). But because rates of first marriage are known to rise and then fall with age (see, e.g., Coale, 1971; Hoem, 1972; Coale and McNeil, 1972) it is desirable that the smoothing technique be able to produce a smooth curve that exhibits substantial curvature. More formally, we require that the smoothing technique be able to reproduce a smooth curve possessing a nonconstant second derivative, which in turn requires that the smoothing procedure be able to vary the degree of smoothing according to local features of the data.

Another practical requirement constraining the choice of a smoothing procedure are the computational resources required to obtain a smooth hazard estimate. For example, although spline and kernel approaches have attractive statistical properties, hazard estimators based on these techniques that have been proposed to date typically require very substantial computational resources when the hazard rate has nonconstant curvature or when sample sizes are very large. This problem is aggravated when attempting to determine confidence intervals because of the difficulties inherent in obtaining confidence intervals from analytic formulas in a setting in which the smoother varies the degree of smoothing in a data dependent manner, as is necessary when attempting to obtain smooth estimates for a hazard rate with nonconstant curvature. In these cases, one might reasonably turn to a computationally intensive nonparametric method like the bootstrap (Efron, 1979, 1982) to obtain confidence intervals. But the use of the bootstrap makes the availability of a fast smoothing algorithm more desirable, particularly for large samples.

A readily available and fast smoothing algorithm that meets these requirements is a variable span, locally linear smoother due to Friedman (1984) and this smoothing algorithm is used in results reported below. Chapter 3 contains a brief description of this smoothing algorithm and the bootstrap procedure employed; for additional details about the smoothing algorithm, see Friedman (1984).

4.3 Exploratory Results

Table 4.2 reports some descriptive statistics for six race-ethnic groups, and white/black breakdowns for four cohorts and educational groups. Some characteristics for these data can be gleaned from the third column, which reports simple within-category percentages. Whites constitute over 80% of the sample while blacks and women of Mexican descent constitute just under 10% and 3% of the sample, respectively. The breakdowns for whites and blacks show that blacks are, on average, somewhat younger and have noticeably lower levels of completed education than whites, although the lower educational attainment is in part due to the younger average age for the black population in these data. Over 75% of blacks were born after 1930 compared to 70% of whites, and 70% of whites had one or more years of college education at the compared to 50%

Table 4.2 Selected statistics for exploratory variables, CPS/W

Variable	Sample size	Percent in category	Raw percent married	Prob. of marrying
Race/Ethnicity				
Whites	50139	82.6%	75.2%	94.5%
Blacks	5928	9.8	58.4	88.4
Mexican	1649	2.7	70.2	96.4
Puerto Rican	385	0.6	57.9	84.6
Cuban	172	0.3	74.4	93.4
Central/South America	249	0.4	63.1	83.0
(Others)	2215	3.6	n.a.	n.a.
Whites, birth cohorts				
1880-1909	3119	6.2	87.2	87.2
1910-1929	12616	25.2	95.2	95.3
1930-1949	15291	30.5	94.6	95.5
1950-1969	19113	38.1	44.5	86.3
Blacks, birth cohorts				
1880-1909	226	3.8	89.4	89.4
1910-1929	1159	19.5	91.7	92.3
1930-1949	1676	28.3	83.5	86.6
1950-1969	2867	48.4	42.3	65.0
Whites, completed education at time of interview				
Grades 0-9	5976	11.9	69.4	93.7
Grades 10-12	9072	18.1	58.7	96.9
Grade 13	20578	41.0	84.1	95.9
Grade 14+	14513	29.0	75.2	91.8
Blacks, completed education at time of interview				
Grades 0-9	1246	21.0	64.5	90.0
Grades 10-12	1640	27.7	46.6	87.7
Grade 13	1864	31.4	62.9	88.1
Grade 14+	1178	19.9	61.0	87.0

of blacks.

The fourth column of Table 4.2 reports the raw within-category percentages of individuals who are married at the time of interview. These values, which report the number married divided the total number in the category and hence do not correct for censoring, differ noticeably from the values in column five of Table 4.2, which report the probability of marriage as derived from the Kaplan-Meier survivor probabilities evaluated at the last observed time for the category and hence do correct for censoring.

Note that the values in columns four and five show substantial agreement for categories in which censoring is minimal, for example, in the pre-1930 birth cohorts. Similarly, the estimates disagree most substantially for those categories in which censoring is heavy. For example, for the 1950 white birth cohort, in which many women have not yet reached the peak marrying age, the raw percentage of women who have married by the time of interview is 44.5%, which is substantially below the raw percentages for the other three white birth cohorts. By contrast, the estimated probability of marriage by age 30 for white women born after 1950 is 86.3%, which agrees much more closely with the corresponding estimates for the other white birth cohorts in Table 4.2, which report the estimated probabilities of marriage by age 50, 70, and 85.

4.3.1.1 Race-ethnic Differences: Survivor Probabilities

Neither the raw percentages in column four nor the estimated probabilities in column five provide much information about how the probability of first marriage for each category varies with age. Figure 4.1 illustrates the Kaplan-Meier survivor curves for whites, blacks, and women of Mexican descent, which are somewhat more informative. There are three curves for each race-ethnic group: a central curve, which gives the estimated Kaplan-Meier survivor probability, and an upper and lower curve, which provide a 10% confidence interval for the estimated survivor probability. These curves report the probability of remaining single; for example, 12% of whites, 25% of blacks, and 12% of women of Mexican descent have not married by age 30; similarly 7% of whites, 14% of blacks, and 5% of women of Mexican descent have not yet married by age 40. But despite clear race-ethnic differences in the proportions who ultimately marry and some difference in the ages at which the survivor probabilities fall most steeply, the qualitative shape of all three survivor curves is quite similar—all curves exhibit a backwards "S" shape. The pointwise confidence intervals in Figure 4.1 show that the survivor probabilities are significantly different for the three populations except for very young ages.

4.3.1.2 Race-ethnic Differences: Integrated Hazards

Figure 4.2 illustrates the Nelson-Aalen estimates of the integrated hazard for the same three race-ethnic groups. Since the survivor probability $S(t)$, integrated hazard $H(t)$ and hazard rate $r(t)$ are closely related according to

$$-\log S(t) = H(t) = \int_0^t r(s)ds$$

the value of $r(t)$ can be determined from the slope of the integrated hazard. Examining the qualitative shape of the integrated hazards in Figure 4.2 we see that, at early ages, the slope of the integrated hazard increases from zero to some maximum positive value slightly after age 20, after which the slope of the integrated hazard slowly declines. From this behavior, we can conclude that the rate of first marriage has a unimodal shape that increases smoothly from zero to a peak sometime after age 20, after which it slowly declines.

4.3.1.3 Race-ethnic Differences: Hazard Rate
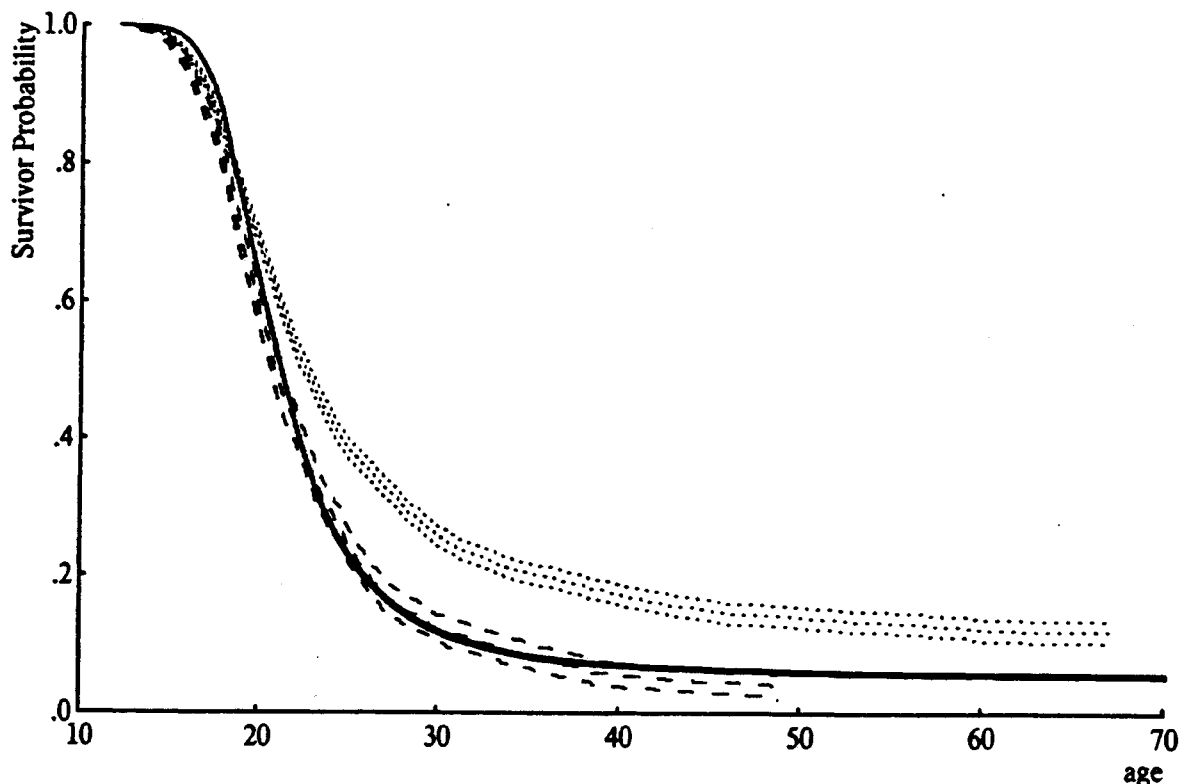
Figure 4.1: Survivor Probabilities: Whites, Blacks, and Mexicans (June 1980 CPS). Three curves for each ethnicity: Kaplan-Meier estimates for survivor probabilities, and pointwise (Greenwood) estimates for the lower and upper 95% confidence interval. Solid curves for whites ($I = 50139$), dotted curves for blacks ($I = 5928$), and dashed curves for women of Mexican descent ($I = 1649$).
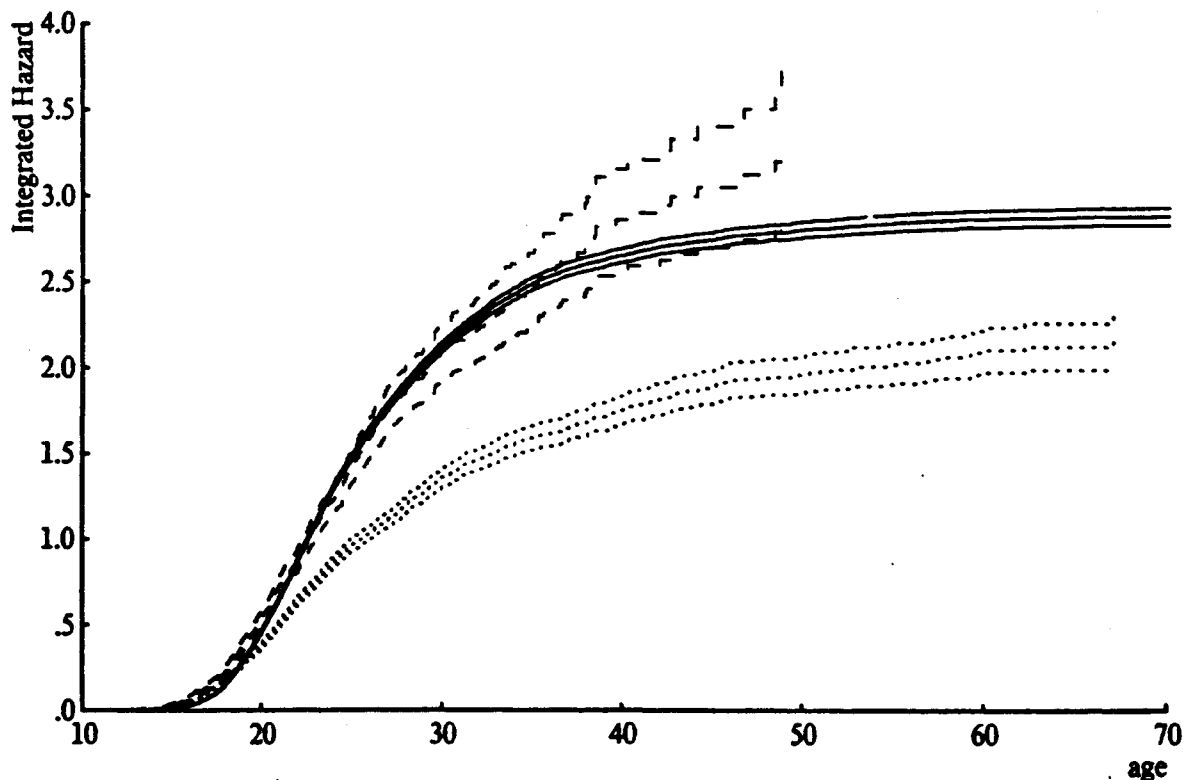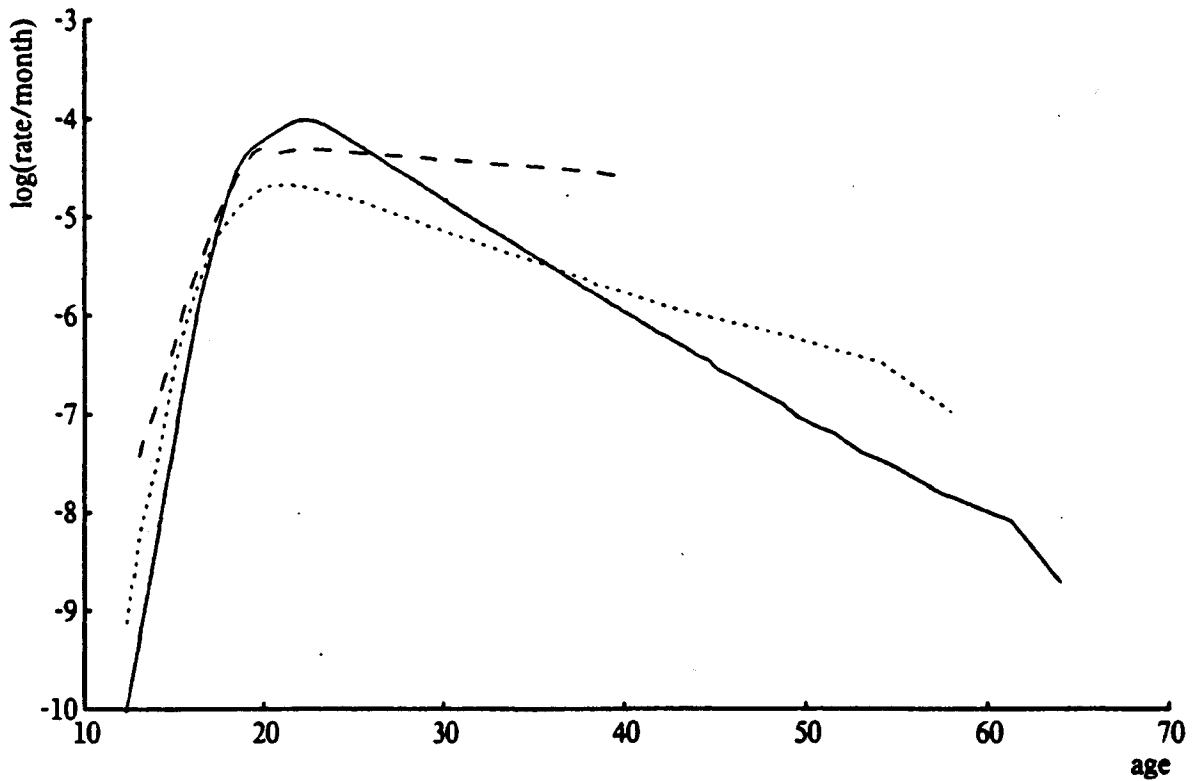

Figure 4.2: Integrated Hazard Estimates: Whites, Blacks, and Mexicans (June 1980 CPS). Three curves for each ethnicity: Nelson-Aalen estimates for integrated hazard, and pointwise estimates for the lower and upper 95% confidence interval. Solid curves for whites ($I = 50139$), dotted curves for blacks ($I = 5928$), and dashed curves for women of Mexican descent ($I = 1649$).

Figure 4.3, which plots the estimates obtained from the smoothed hazard estimator introduced in Chapter 3, provides far more information by which to judge the similarities and differences in the rate of first marriage between the three race-ethnic groups. The rate of first marriage for white women is characterized by a smooth rise in the rate at early ages to a peak at age 22 followed by smooth decline. The rate of first marriage for black women is characterized by substantially lower and more rounded peak at age 21, followed by a gradual decline in the rate for later ages. Note that despite the substantially lower peak rate of marriage for black women, which is close to half that of white women, the black rate of marriage is slightly higher in absolute terms after age 40. The rate of first marriage for women of Mexican descent is characterized by a smooth rise in the rate at early ages to a peak at age 20, followed by an extremely gradual decline. All three populations exhibit quite different peak rates of marriage, with 20 per 100 per month for white women, 15 per 100 per month for women of Mexican descent, and 10 per 100 per month for black women. Similarly, the rates of marriage exhibit quite different shapes at later ages, with little decline for women of Mexican descent and relatively greater decline for white and black women. Interestingly, differences in the peak age of marriage are less striking in Figure 4.3 with the péak age of marriage occurring between the ages of 21 and 23 for all three populations.

Figure 4.4 illustrates the natural logarithm of the smoothed hazard rate. There are several reasons why looking at the logarithm of the rate might prove helpful when examining the effects of an individual attribute such as such race-ethnicity. First, the estimator introduced in Chapter 3 applies a smoothing procedure to the estimates $\log \hat{\rho}_j$ to obtain smoothed estimates $\log \tilde{\rho}_j$. Thus an examination of the $\log \tilde{\rho}_j$ may prove useful in determining how features of the smoothed hazard rate estimates $\tilde{\rho}_j$ were obtained from the rough estimates $\log \hat{\rho}_j$. Second, it is standard practice to incorporate covariates into parametric models of the rate in a log-linear manner, for example, in the so-called proportional hazard model

$$\log r_i(t) = \beta x_i$$

where x_i is a vector of covariates for individual i, and β is a vector of parameters. Thus observing the effect of a covariate on the log rate provides an informal means for assessing the usefulness of a proportional model and, if violations of proportionality are found, may suggest simple alternatives to the proportional hazard model. Lastly, because we suspect that rates of first marriage may exhibit a nonlinear pattern of age dependence, it seems reasonable to examine simple transformations of the rate that may reveal more linear patterns of age dependence.

The smoothed log hazard estimates in Figure 4.4, like the smoothed hazard estimates in Figure 4.3, are quite informative. A striking feature of the smoothed log hazard estimates not entirely apparent in Figures 4.1-4.3 is that the rate of first marriage appears to be governed by a strongly regular process despite clear variations across race-ethnic groups. The qualitative shape of the hazard rate for all three race-ethnic groups appears to follow a more or less linear ascent to a peak around age 20 followed by a more or less linear decline from the peak. Figure 4.4,
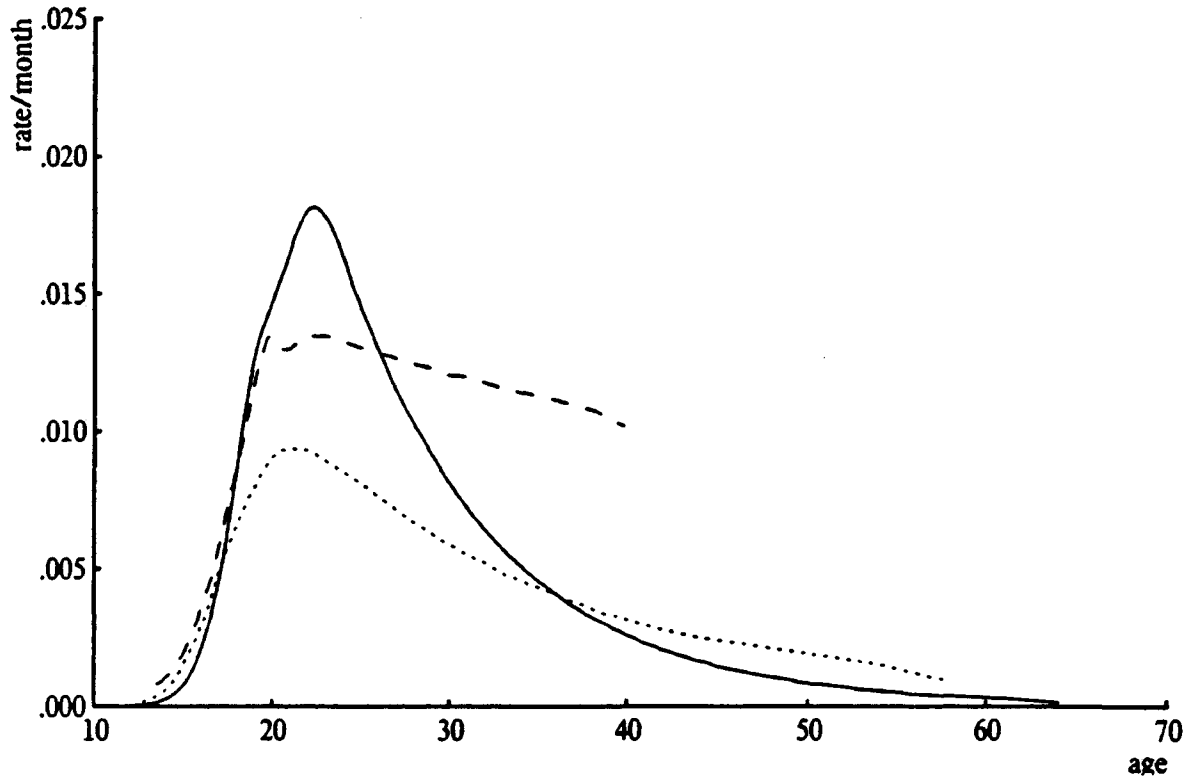
Figure 4.3: Smoothed Hazard Estimates: Whites, Blacks, and Mexicans (June 1980 CPS). Solid curve for whites $(I = 50139)$, dotted curve for blacks $(I = 5928)$, and dashed curve for women of Mexican descent $(I = 1649)$.

Figure 4.4: Smoothed Log Hazard Estimates: Whites, Blacks, and Mexicans (June 1980 CPS). Solid curve for whites $(I = 50139)$, dotted curve for blacks $(I = 5928)$, and dashed curve for women of Mexican descent $(I = 1649)$.

like Figure 4.3, provides evidence of substantial variation in the height of the peak of the rate of marriage for the three race-ethnic groups. Figure 4.4 also provides evidence that the differential decline of the rate after the peak age of marriage between race-ethnic groups that was found in Figure 4.3 can be described by different linear slopes that characterize the log rate for these groups. The logarithm of the hazard rate of first marriage for white women is characterized by a relatively steep ascent in the rate to age 22, followed by a relatively rapid decline. In contrast, the log hazard rate for black women is characterized both by a substantially lower peak rate of marriage at age 21 *and* a slower decline for later ages. Lastly, the log hazard for women of Mexican descent is characterized by a peak rate of marriage at age 20 whose value lies between those for whites and blacks, and a extremely slow decline in the rate for later ages.

4.3.1.4 Race-ethnic Differences: Evaluation of Smoothing Methods

From a methodological standpoint, the smoothed hazard estimator introduced in Chapter 3 appears to be highly successful for these data. Figures 4.5, 4.6, and 4.7 illustrate the rough estimates $\log \hat{\rho}_j$ obtained from the actuarial/incidence estimator in Equation (3.12) and the smoothed estimates $\log \tilde{\rho}_j$ obtained by passing the variable span, local linear smoothing algorithm due to Friedman through the rough estimates. In Figure 4.5, the "rough" estimates $\log \hat{\rho}_j$ themselves virtually describe a smooth curve, particularly between the ages of 12 to 30, in part because of the extremely large sample of white women available in these data. The rough estimates in Figures 4.6 and 4.7, which are based on a considerably smaller samples of black women and women of Mexican descent, show substantially greater variation. Nevertheless, the smoothed curves in the three figures appear to be fitting the $\log \hat{\rho}_j$ in an intuitively reasonable way.

These data in fact provide a difficult smoothing problem. Recall that the intent of smoothing is to recover a smooth function from "noisy" data. Although the majority of smoothing procedures perform well for data generated by a hazard rate $r(t)$ that exhibits constant curvature (that is, for functions $r(t)$ that have a second derivative that is constant or that changes slowly with age), it is far more difficult to recover a smooth estimate for a hazard rate that exhibits highly nonconstant curvature. But Figures 4.5-4.7 strongly suggest that the slope of logarithm of the hazard rate for first marriage changes substantially with age and hence that the underlying log hazard has a second derivative that changes substantially with age.[5] The variable span smoothing algorithm adapts to the nonconstant curvature in these data by choosing small spans in the region where curvature is highest and larger spans where curvature is low. In contrast, note that the choice of a constant span would undersmooth the portions of the log hazard with low curvature and oversmooth the portions of the log hazard with high curvature. These

---

[5] An informal analysis of the smoothed hazard estimates in Figures 4.5-4.7 shows that the second derivative for the log hazard of marriage takes predominately negative values, which is consistent with the observed concavity of the log hazard estimates. The second derivative takes values close to zero for the portions of the log hazard curves that appear linear, and assumes increasingly negative values near the peak ages of marriage, with the most negative values occurring just prior to the peak age of marriage.
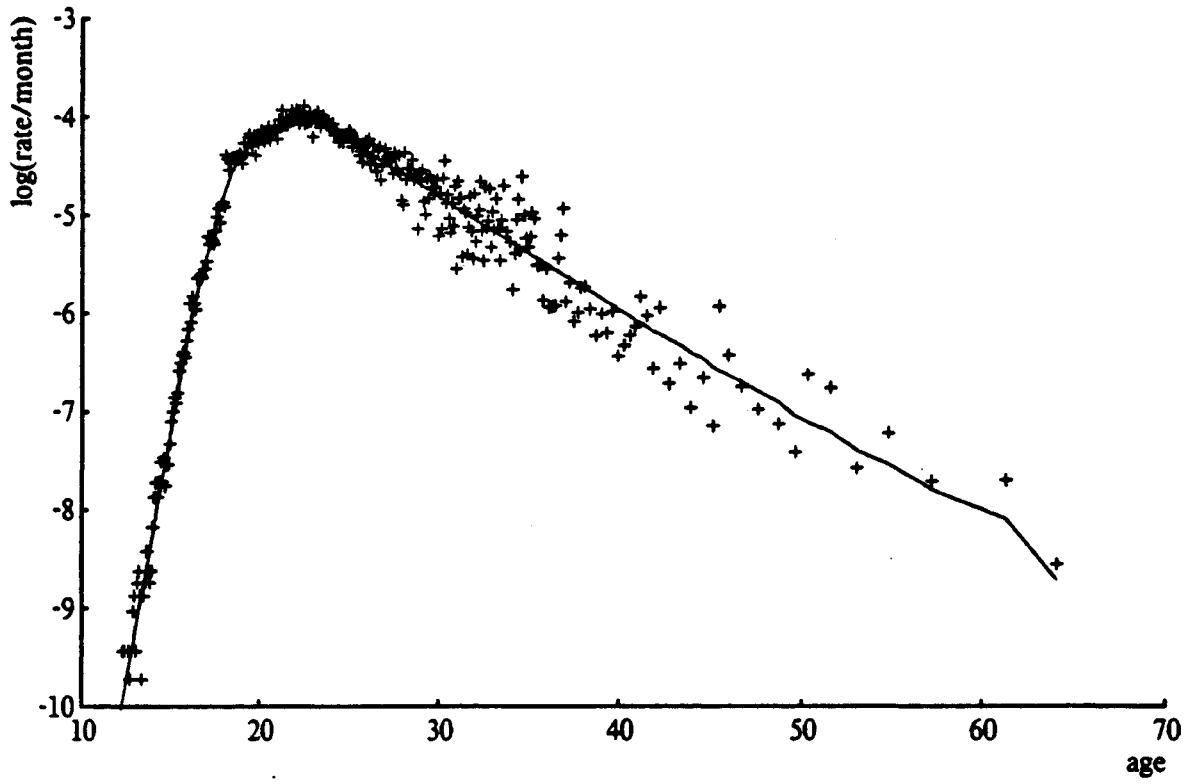
Figure 4.5: Empirical and Smoothed Log Hazard Estimates: Whites (June 1980 CPS).
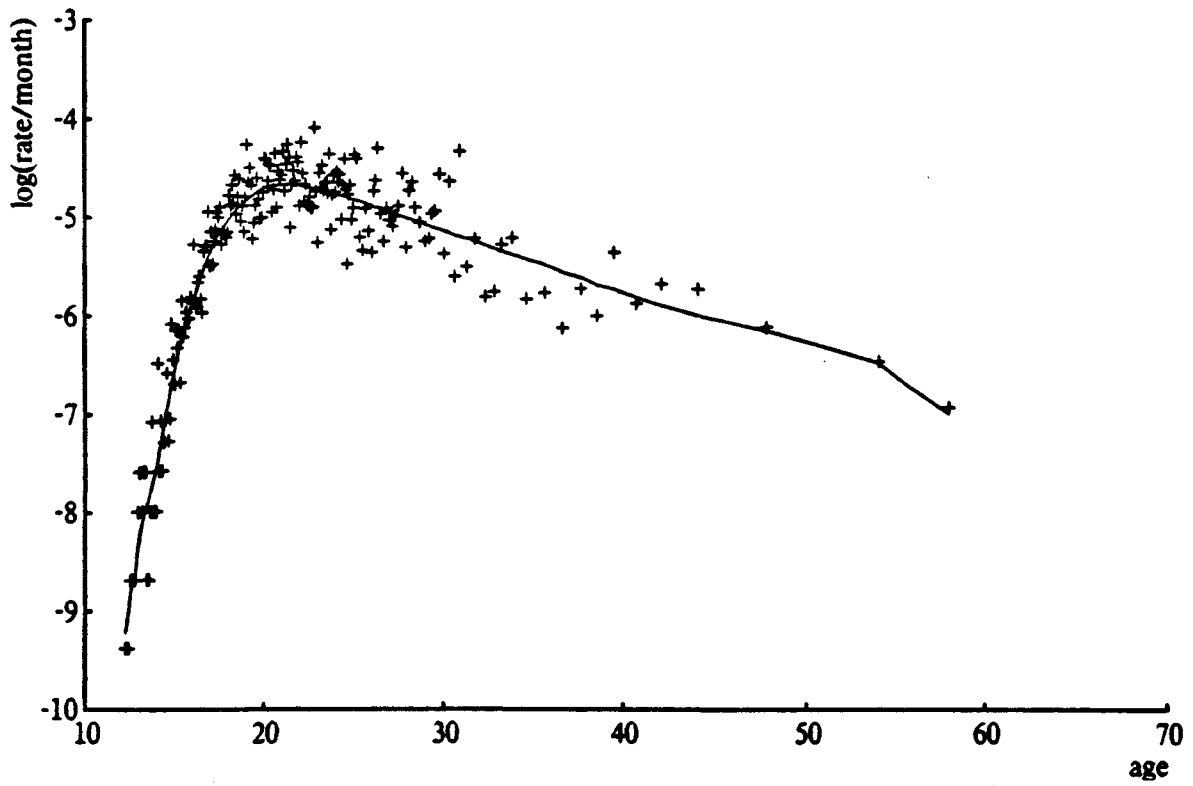
Figure 4.6: Empirical and Smoothed Log Hazard Estimates: Blacks (June 1980 CPS).

considerations underscore the importance of utilizing a smoothing procedure that performs well for problems in which we can anticipate nonconstant curvature of the underlying hazard.

4.3.1.5 Bootstrapping the Log Hazard Estimates for Race-Ethnic Differences

This section presents bootstrap results that assess the quality (bias and variability) of the smoothed log hazard estimates presented in Figure 4.4. At face value, the relatively large sample sizes for each race-ethnic group might seem to guarantee that the log hazard curves have low variability and bias, and hence that observed differences are statistically significant. Nevertheless there are several considerations that make an assessment of bias and variability important. First, although the sample sizes available for the three race-ethnic are quite large, the statistical efficiency of the smoothed hazard estimator may be quite low compared to the use of parametric estimators when we have knowledge of the true parametric model. Secondly, smoothing procedures by their nature introduce bias and variability into estimates. In general, smoothing procedures must strike a compromise in the degree of smoothing used to obtain estimates: oversmoothing the data will introduce bias into the estimates while undersmoothing will produce estimates that have more variability than is optimal. The issue of bias and variability is further complicated because the variable span smoothing procedure varies the degree of smoothing locally in a data dependent manner. Lastly, I argued in Chapter 3 that one could more reliably smooth on the estimates $\log \hat{\rho}_j$ since some heuristic arguments suggested that the sampling distribution of these estimates was close to a Gaussian distribution with zero mean for intervals containing a sufficient number of observed events. But I also noted that the various approximations used in the heuristic argument become increasingly inaccurate as the number of observed events in the time interval $[t_{j-1}, t_j)$ decreases. Thus, these considerations suggest that large sample sizes may not automatically guarantee significant results.

Although bootstrapping requires considerable computational resources to obtain results, it is possible to realize substantial savings in bootstrapping the smoothed hazard estimator. The savings are possible because once the rough estimates $\hat{\rho}_j$ have been computed for the J time intervals $[t_{j-1}, t_j)$, the bootstrap procedure described in Chapter 3 can be applied to the number of individuals who experience events or who are censored censored within each interval. This reduces the computational complexity of the bootstrap procedure from a problem of order I, the number of individuals in the sample, to a problem of order J, the number of discrete time intervals. In practice the savings can be considerable. For example, computing the original rough estimates $\hat{\rho}_j$ for the 3 race-ethnic groups on an IBM 4381 running VM/CMS took approximately two cpu minutes. Computing 250 bootstrap iterations for these data took approximately 45 cpu minutes as opposed to $250 \times 2 = 500$ cpu minutes that a more inefficient bootstrap algorithm might require.

4.3.1.6 Bootstrap Results for Race-Ethnic Differences

Figure 4.8 presents bootstrap bias-corrected hazard estimates and 90% confidence intervals based on 250 bootstrap iterations, and compares these results to the original smooth hazard
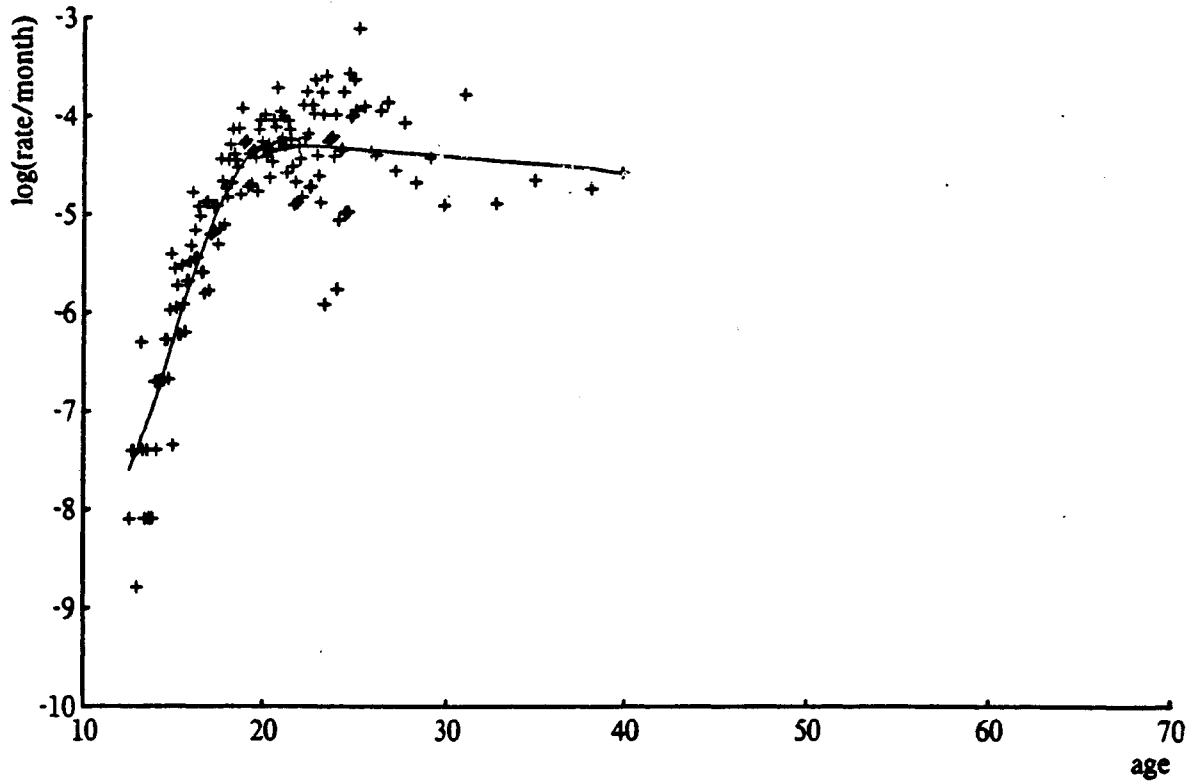
Figure 4.7: Empirical and Smoothed Log Hazard Estimates: Women of Mexican Descent (June 1980 CPS).
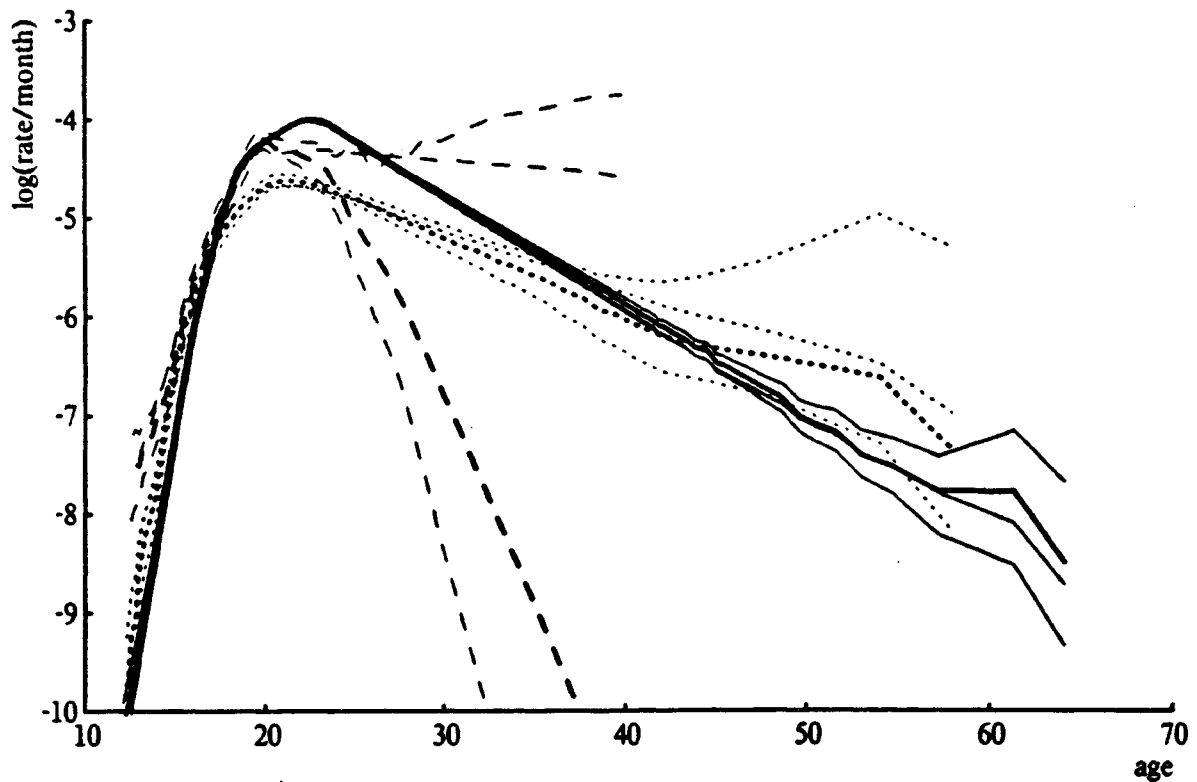

Figure 4.8: Bootstrap Results for Smoothed Log Hazard Estimates: Whites, Blacks, and Mexicans (June 1980 CPS). Four curves for each ethnicity: Original smoothed log hazard estimate (light line), bias-corrected smoothed log hazard, lower and upper bootstrap 90% pointwise confidence intervals. Solid curves for whites ($I = 50139$), dotted curves for blacks ($I = 5928$), and dashed curves for women of Mexican descent ($I = 1649$).

estimates. (For a description of the bootstrap procedure used to obtain bias-corrected estimates and intervals, see Section 3.4.3 in Chapter 3.) Note that the bootstrap estimates of the bias-corrected log hazard rate for the three race-ethnic groups reveal that bias increases at later ages, and becomes noticeably large for smaller samples and when the number of events becomes sparse.

Where estimated bias is small, the 90% bootstrap confidence intervals are, for the most part, quite close to both the original smoothed estimates and the bias corrected bootstrap estimates of the log hazard. But where estimated bias is large, for example, the bootstrap estimates of bias after age 25 of the log hazard for women of Mexican descent, the 90% bootstrap confidence intervals become extremely wide. This suggests that in small samples and for sparse events, the estimation procedure introduced in Chapter 3 is subject to substantial bias, and that bias of these estimates tends to dominate the variability of the estimates; conversely, in large samples or for time intervals for which there exist a large number of events, the estimation procedure appears to produce estimates with low estimated bias and variability. Lastly, it is interesting to note that the width of pointwise confidence intervals for the smooth hazard estimator do not increase monotonically with age, as do the standard confidence intervals for the survivor and integrated hazard functions. In this respect, the confidence intervals behave in a manner more like confidence intervals for a histogram estimate, which are dependent on the observed frequency within bins. Generally speaking, the confidence intervals in Figure 4.8 are relatively wider at very early ages and just before the peak age at marriage where the curvature of the log hazard estimate is greatest.[6]

Although the pointwise confidence intervals in Figure 4.8 are quite informative, they do not speak directly to some aspects of the log hazard that are of substantive interest. The results in Section 4.3.1.3 suggested that three distinct aspects of the log hazard rate of marriage vary in a systematic way with race-ethnicity. First, we observed differences in the largest value taken by the log hazard at the peak age of marriage. Second, we observed variations in age at which the peak of the log hazard occurs. Lastly, we observed that the log hazard appears to undergo a linear decline at later ages with variation in the slope of the linear decline. Assessing the variability and significance of these aspects of the log hazard are considerably more difficult than forming simple confidence intervals and, indeed, there is some evidence that computing bootstrap intervals for the mode-like problem for the age at which the log hazard reaches a maximum is a difficult problem (Romano, 1986). The remainder of this section provides an informal assessment of the significance of the first and third observed regularities: the observed differences in the largest value attained by the log hazard and the slope of the hazard for later ages.

---

[6] The relatively wide intervals at very early ages can be attributed to the relative scarcity of events at these ages. And as previously noted, the problem of estimating a smooth curve when the data exhibit high curvature is a difficult one, which suggests that one might reasonably expect estimates to exhibit somewhat greater variability in areas in which the curve has high curvature.

Because the time intervals around the peak ages of marriage have large numbers of events, the bootstrap pointwise confidence intervals are reasonably narrow about the bias corrected bootstrap log hazard estimates. Nevertheless, it is not possible to directly determine whether the value of log hazard estimates at the peak age of marriage are significantly different from one another by a simple inspection of Figure 4.8 because of the differences in the age at which the peak occurs. However the following figures provide some indication that the differences are significant. Using the 95% percentile points from the bootstrap results (a significance level slightly higher than that used in plotting confidence intervals Figure 4.8) yields a peak value of -4.00 for the log hazard at age 22.5 with a 95% confidence interval of (-4.03,-3.98) for white women, -4.62 at age 21.5 with a 95% confidence interval of (-4.68,-4.55) for black women, and -4.21 at age 20.0 with a 95% confidence interval of (-4.32,-4.07) for women of Mexican descent. Since the 95% bootstrap confidence intervals do not overlap, these results provide preliminary evidence that we may reject the null hypothesis that the observed differences in the values of the log hazard estimates at the peak age of marriage for the three race-ethnic groups are due to chance, with the proviso that these preliminary results do not control for other aspects of population heterogeneity and under the assumption that we may reasonably compare the peak values of different log hazard curves even if they occur at different ages.[7]

Assessing differences in the slope of the log hazard at later ages presents a more difficult problem and to simplify the discussion, I omit a comparison of the log hazard estimates for women of Mexican descent because the large bootstrap estimates of bias cast doubt on the reliability of both the original smooth hazard estimates and the bias corrected bootstrap estimates for ages after 25. To further simplify matters, I utilize a crude estimate of the slopes of the two log hazard curves obtained by calculating the slope of the line connecting the estimated log hazards at ages 25 and 40. Then a conservative lower (upper) confidence interval for the slope estimate can be similarly obtained by connecting the upper (lower) confidence point at age 25 with the lower (upper) confidence point at age 40. This crude procedure yields an estimated slope of .110 with a 95% confidence interval of (.108,.113) for the log hazard for white women and an estimated slope of .081 with a 95% confidence interval of (.055,.103) for the log hazard for black women.[8] Because these crude 95% confidence intervals do not overlap, the results suggest that between the ages of 25 and 40, we can provisionally reject the null hypothesis that the slopes for the log hazard rates for white and black women are equal, again with the proviso that these preliminary results do not control for other aspects of population heterogeneity and under the assumption that we may reasonably model the behavior of the log hazard curves between

---

[7] More formally, I assume that the log hazard functions for one group is related to the second by a translation in time of the first (see, e.g., Cox and Oakes, 1984).

[8] Choosing the upper age point to be 45 yields similar results with an estimated slope of .109 with a 95% confidence interval of (.105,.113) for the log hazard for white women and an estimated slope of .079 with a 95% confidence interval of (.040,.097) for the log hazard for black women. The curve for black women contains only a few time points beyond age 45 and as a result the bias of the estimates increases for these ages. Because of this, 45 represents a reasonable upper age limit for the purposes of comparing these curves.

the ages of 25 and 40 by a linear function.

4.3.2 Birth Cohorts: Black/White Differences

Demographers often examine age, period, and cohort effects in their analyses, and in an influential paper, Ryder (1965) has made these issues standard among sociologists as well. Section 4.3.2.1 presents bias corrected bootstrap log hazard estimates and confidence intervals for four broadly defined birth cohorts of black and white women born between 1880-1909, 1910-1929, 1930-1949, and after 1950. Section 4.3.2.2 uses the bootstrap estimates in an informal assessment of age, period, and cohort effects. Preliminary results show that a black/white effect and age and period effects are needed to account for the observed peak values for three of four of the black and white cohorts; a cohort effect only appears necessary to account for black/white differences for women who reached adulthood during the Depression.

4.3.2.1 Birth Cohorts: Bootstrap Results

Figures 4.9a and 4.9b present bias-corrected bootstrap estimates and 90% confidence intervals for the log hazard rate for four broadly defined birth cohorts of white women; Figures 4.10a and 4.10b present the corresponding estimates for blacks. Note that disaggregating results by birth cohort necessarily limits the range of ages for the abscissa of each log hazard plot. For example, no log hazard estimates are possible after 30.5 for women in the 1950+ cohort since no women in this group was older than 30 years and 6 months in June 1980, the date of the interview.

These results also clearly demonstrate the increased variability of estimates for ages near the top of the range for the birth cohort. As was the case for the bootstrap estimates for the three race-ethnic groups, the bias and variability of estimates for later ages tends to increase in inverse proportion to the sample size, but it is interesting to note that the bias and variability of estimates at extreme ages is evident for all birth cohorts and is particularly severe for the most recent birth cohorts, which have large numbers of women ($I = 19113$ and 2867 for whites and blacks, respectively).

Note that the log hazards for all birth cohorts have relatively broad peaks in the 20-25 age interval; for women born before 1950 the log hazards decline in a relatively linear fashion after age 25, in a manner that is in qualitative agreement with the regularities noted for the log hazards estimated for the race-ethnic groups in Section 4.3.1. If this general pattern is true for the most recent cohorts, then estimating a linear declining slope from relatively sparse data in the 25-30 age range can be anticipated to be a difficult problem, which would account for the bias and variability observed in the estimates for these cohorts. That is, in order to closely follow the high curvature of the log hazard curves for ages between 18 and 25, a variable span smoothing procedure must necessarily choose a small span for these ages. If sufficient numbers of events are observed for a wide range of ages after 25, then a variable span smoothing procedure can choose large smoothing spans in estimating the log hazard for ages greater than 25. However, if the data are sparse between 25 and 30 a variable span smoothing procedure might retain the small span values used in the age ranges between 18 and 25. These results suggest that bias
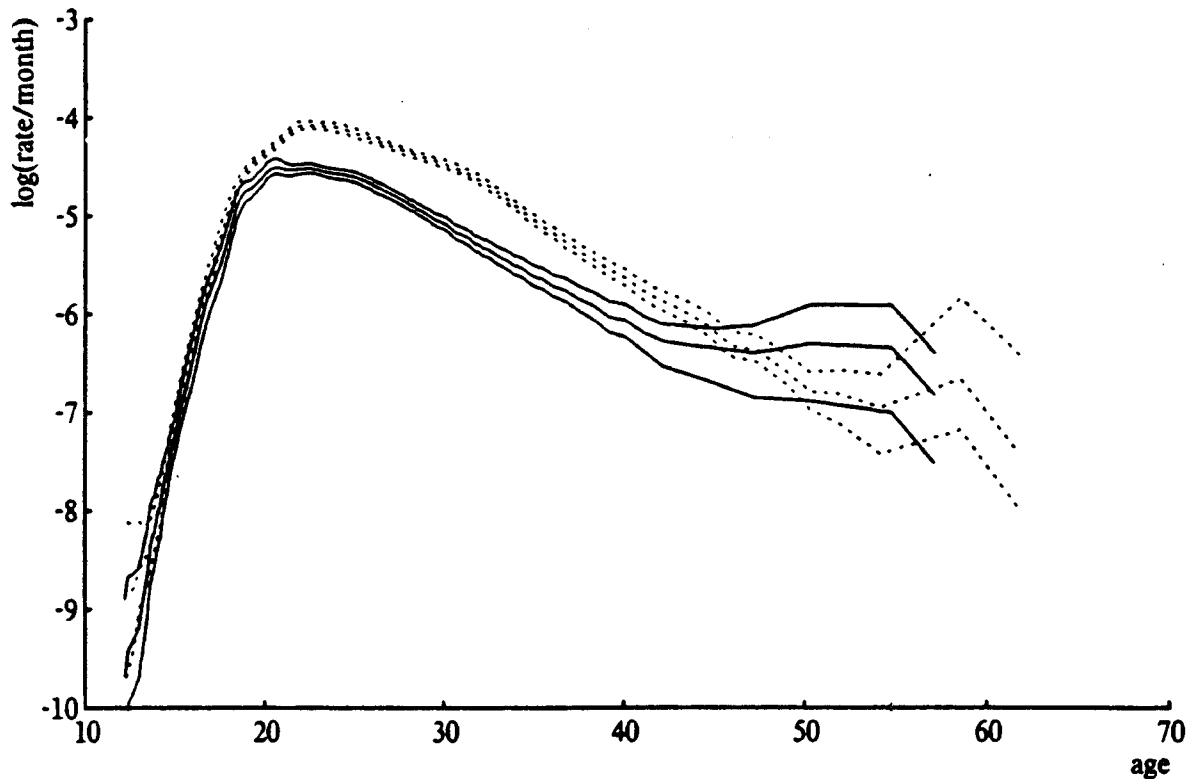
Figure 4.9a: Bootstrap Results for Smoothed Log Hazard Estimates: White Birth Cohorts (June 1980 CPS). Three curves for each cohort: Bias-corrected smoothed log hazard, lower, and upper bootstrap 90% pointwise confidence intervals. Solid curves for women born 1880-1909 ($I = 3119$) and dotted curves for women born 1910-1929 ($I = 12616$).
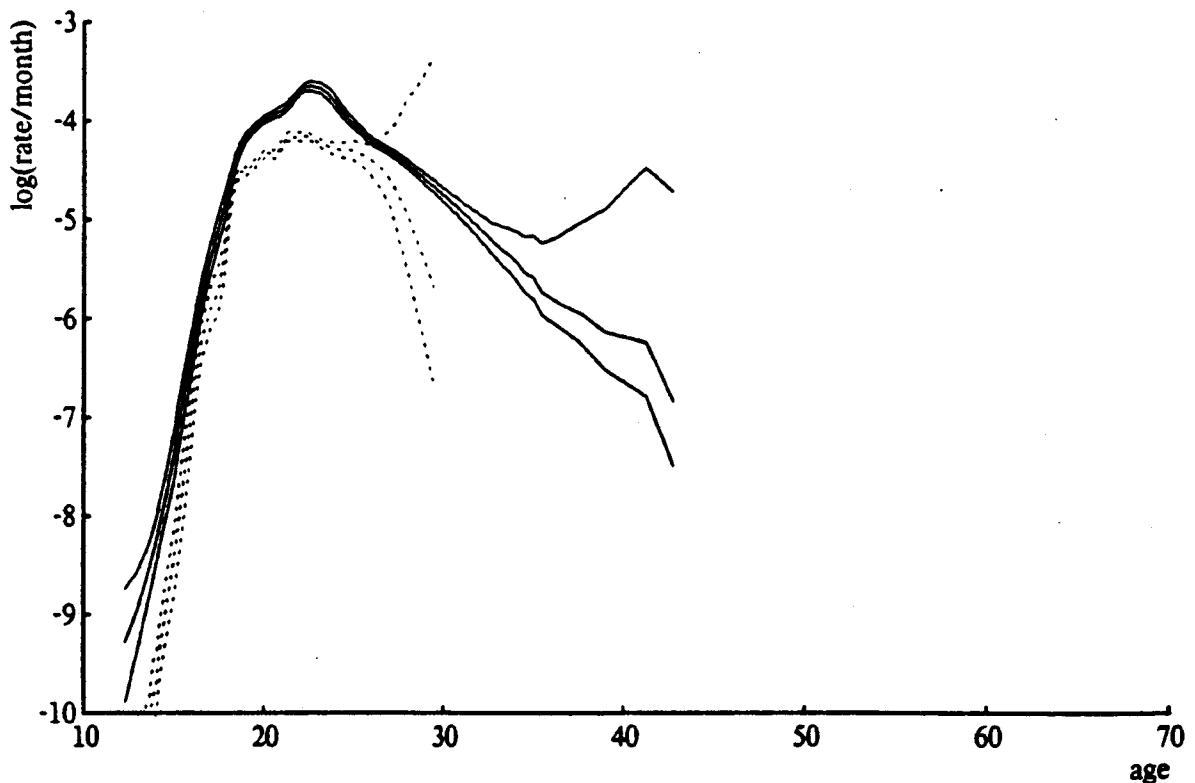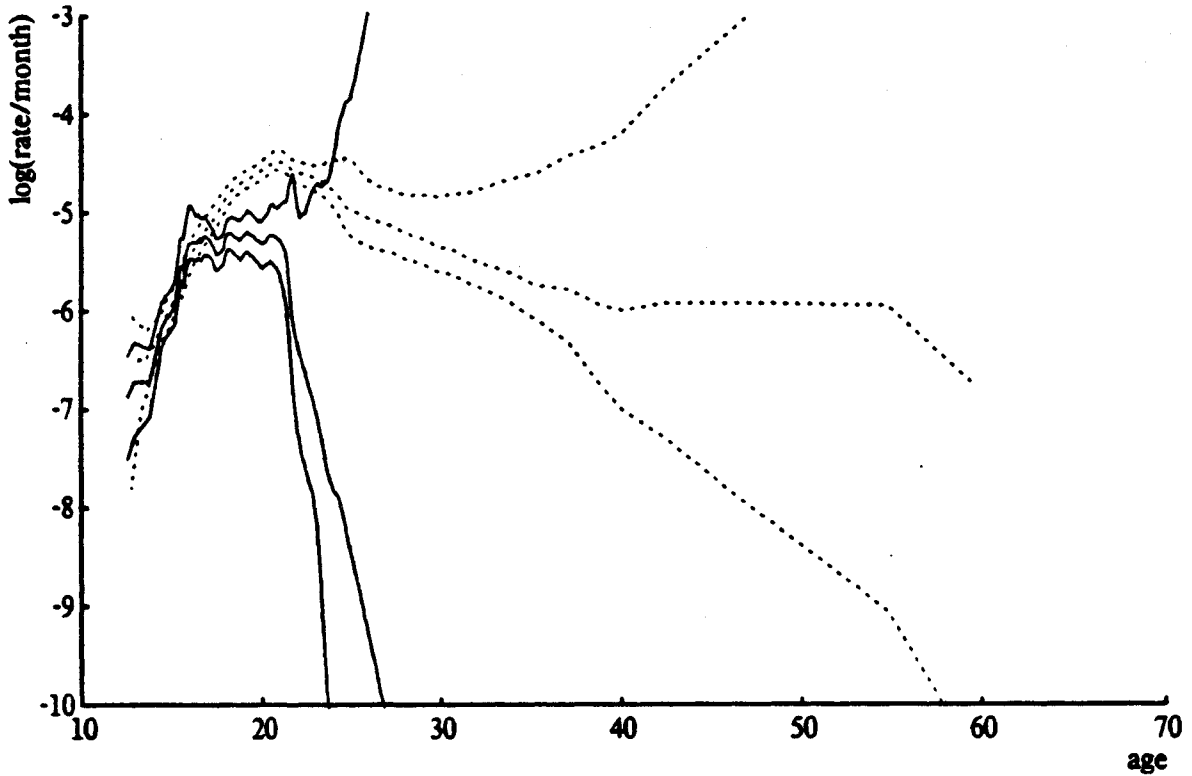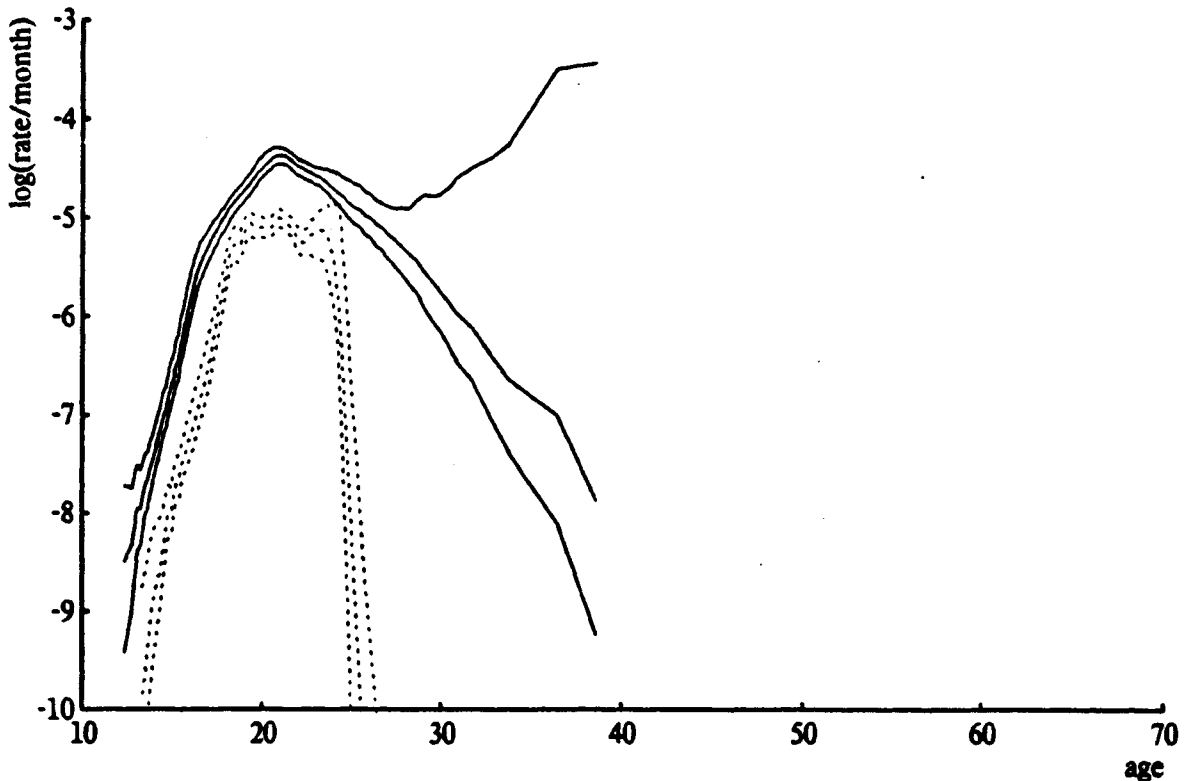

Figure 4.9b: Bootstrap Results for Smoothed Log Hazard Estimates: White Birth Cohorts (June 1980 CPS). Three curves for each cohort: Bias-corrected smoothed log hazard, lower, and upper bootstrap 90% pointwise confidence intervals. Solid curves for women born 1930-1949 ($I = 15291$) and dotted curves for women born after 1950 ($I = 19113$).

Figure 4.10a: Bootstrap Results for Smoothed Log Hazard Estimates: Black Birth Cohorts (June 1980 CPS). Three curves for each cohort: Bias-corrected smoothed log hazard, lower, and upper bootstrap 90% pointwise confidence intervals. Solid curves for women born 1880-1909 ($I = 226$) and dotted curves for women born 1910-1929 ($I = 1159$).

Figure 4.10b: Bootstrap Results for Smoothed Log Hazard Estimates: Black Birth Cohorts (June 1980 CPS). Three curves for each cohort: Bias-corrected smoothed log hazard, lower, and upper bootstrap 90% pointwise confidence intervals. Solid curves for women born 1930-1949 ($I = 1676$) and dotted curves for women born after 1950 ($I = 2867$).

and variability are affected in a complicated way not only by the overall sample size, but also by an interaction of the relative curvature of the underlying log hazard in an age range and the number of events observed for age intervals in this range.

Overall, there is a general historical trend for both white and black populations of an increase in the peak log rate of marriage from 1880 to 1950, followed by a decline in the peak log rate after 1950. It also appears that black log rates of marriage decline slightly less quickly with age than do white log rates, although these differences are far less marked than observed in Section 4.3.1. For both white and black populations, the 1930-1949 cohort shows evidence of both an unusually high peak in the log hazard and an unusually steep downward slope after this peak.

4.3.2.2 Birth Cohorts: Assessing Cohort Effects

In a review of the literature of age, period, and cohort models, Hobcroft, Menken, and Preston (1982) question the usefulness of models incorporating cohort effects and ask "Do age patterns, for individuals who have the timing of an event (such as their birth or marriage) in common, respond only to period influences, so that the experience of a cohort can be described completely by age effects and the effects of the periods its members live through, or do they respond to additional, cohort influences as well?" Building on this incisive question, we might suppose that, in the context of accounting for observed trends in the peak magnitudes of the log hazard rates for black and white cohorts, an adequate age-period model is one that accounts for observed differences without recourse to a cohort effect, where the period effect is an effect that holds equally for black and white populations, net of an overall black/white effect.

This suggests that a conservative test for assessing the need for a cohort effect would be to attempt to predict peak values for the log hazard of black cohorts from the observed peak values for the hazard log hazard rates of the corresponding white cohorts by means of a single black/white effect, where the observed white peak values of the log hazard provide crude baseline estimates for the historical variation in the peak rates of marriage and the black/white effect accounts for the relative difference in the peak rates of marriage between blacks and whites. Then for the 1880 cohorts, the peak values are -4.509 at age 20.6 for whites and -5.187 at age 19.1 for blacks, for a difference of .678; for the 1910 cohorts, the peak values are -4.087 at age 22.0 for whites and -4.470 at age 20.9 for blacks, for a difference of .383; for the 1930 cohorts, the peak values are -3.647 at age 22.6 for whites and -4.371 at age 21.1 for blacks, for a difference of .724; and for the 1950 cohorts, the peak values are -4.152 at age 22.5 for whites and -5.001 at age 21.0 for blacks, for a difference of .849.

The average of the black/white differences (.66) provides a crude estimate of a black/white effect, and predicting the peak values for the log hazard rate for black cohorts by subtracting .672 from the observed white peak values produces estimates that lie outside the 95% boostrap confidence intervals (not reported) for black peak values for two of the four black cohorts. However, further inspection suggests that the observed black/white difference for the 1910 cohort

is unusually small relative to the differences for other cohorts. Then using the average of the black/white difference for the three remaining cohorts (.75) produces predicted peak magnitudes for the 1880, 1910, and 1950 black cohorts that lie within their respective 95% bootstrap confidence intervals.[9]

These results suggest that, for three of the four cohorts, the observed differences in the peak magnitudes of the log hazard rates for black and white cohorts can be adequately described in terms of a black/white effect and a period effect, with the usual proviso that these preliminary results do not control for other aspects of population heterogeneity and under the assumption that we may reasonably compare the peak magnitudes of different log hazard curves even if they occur at different ages. However, black/white differences for women born between 1910 and 1929 were smaller than for women born in other years, and the peak value for the log hazard for black women born between 1910 and 1929 was relatively higher than would be expected from the trends based on the experiences of other cohorts. This suggests that white and black peak rates of marriage were relatively closer for women who reached adulthood during the Depression era than in other eras.

4.3.3 Completed Education

Figures 4.11a,b and 4.12a,b present bias-corrected estimates and 90% confidence intervals for the log hazards for four educational levels (grades 0-9, 10-12, 13, and 14+); Figures 4.11a and 4.11b present estimates for whites and Figures 4.12a and 4.12b present estimates for blacks. Many of the same overall patterns observed in previous figures are also replicated here. The overall shape of the log hazard follows a pattern characterized by a linear ascent of the log rate to a peak after which the log rate declines in a linear fashion. Variability increases at later ages and bias (not reported in Figures 4.11 and 4.12) is considerably greater at later ages for blacks.

For both white and black populations, the effect of more education tends to shift the peak of the log hazard curve to later ages, although the observed age shift is small and the tendency is violated for some of the educational groups. For whites, the peak rate of marriage is -4.23 at age 21.3 with a 95% confidence interval of (-4.28,-4.17) for women with 0-9 years of education; -3.86 at age 18.6 with a 95% confidence interval of (-3.91,-3.79) for women with 10-12 years of education; -3.847 at age 21.8 with a 95% confidence interval of (-3.88,-3.81) for women with 13 years of education; and -3.990 at age 22.8 with a 95% confidence interval of (-4.03,-3.95) for women with 14+ years of education. For blacks, the figures are -4.553 at age 18.1[10] with a 95% confidence interval of (-4.71,-4.37) for women with 0-9 years of education; -4.674 at age 20.4 with a 95% confidence interval of (-4.80,-4.54) for women with 10-12 years of education; -4.464 at age 19.9 with a 95% confidence interval of (-4.58,-4.34) for grade 13 years of education; and

---

[9] Using a slightly higher value, .80, for the black/white effect produces predicted black peak magnitudes for the three black cohorts that lie within the narrower 80% confidence intervals.

[10] The estimates for the 1880-1909 black cohort actually attain their highest values after age 35. However the estimated bias and variability of estimates after age 35 is extremely large, and I have chosen to report the maximum value attained by the log rate for ages less than 35 for this cohort.
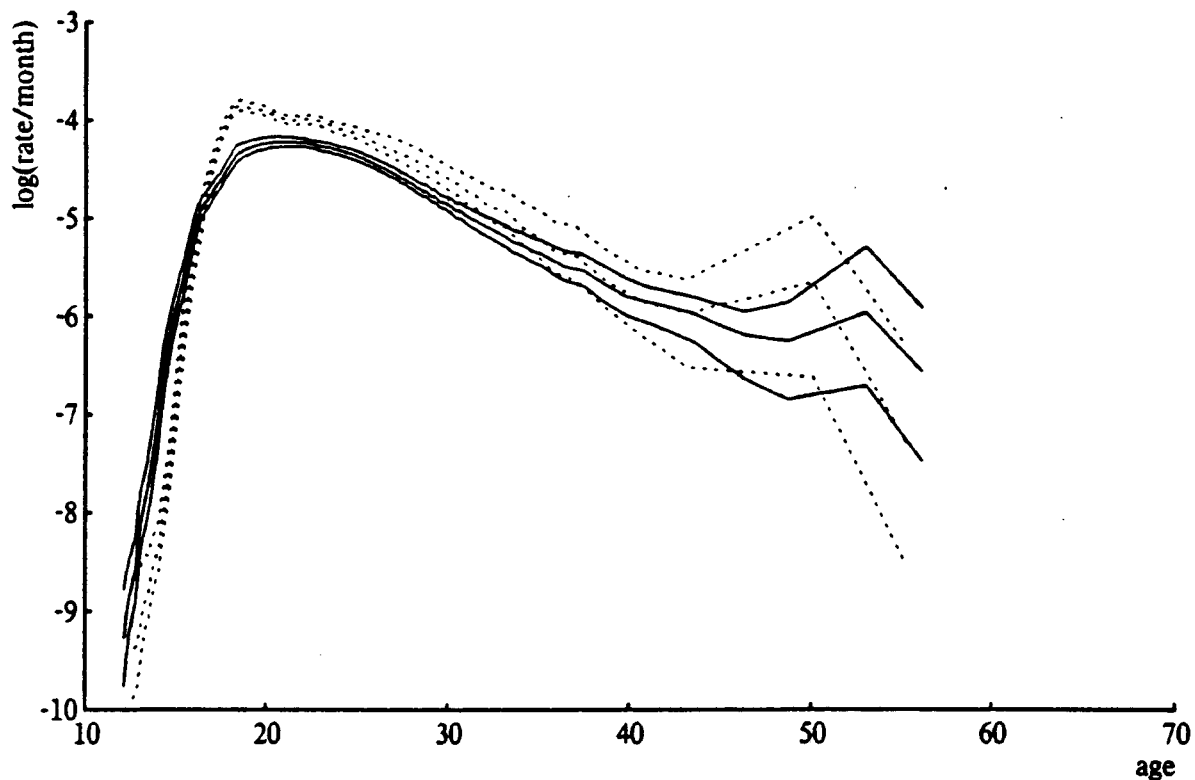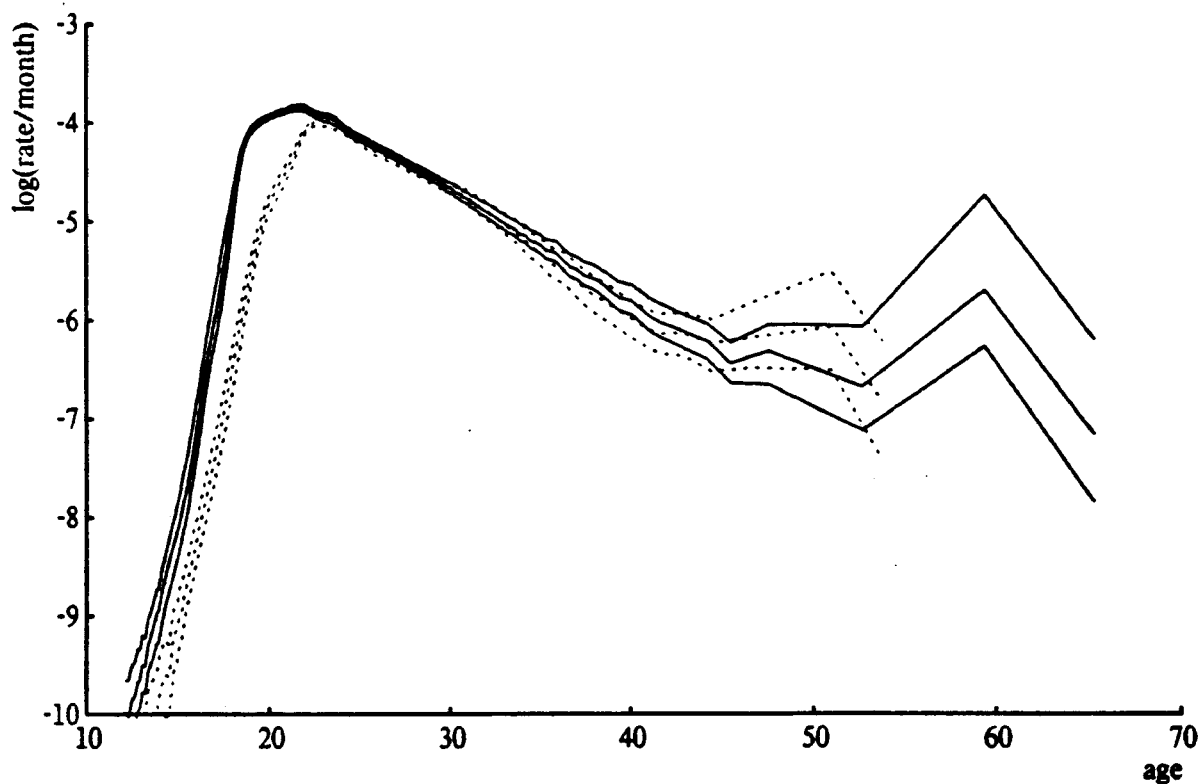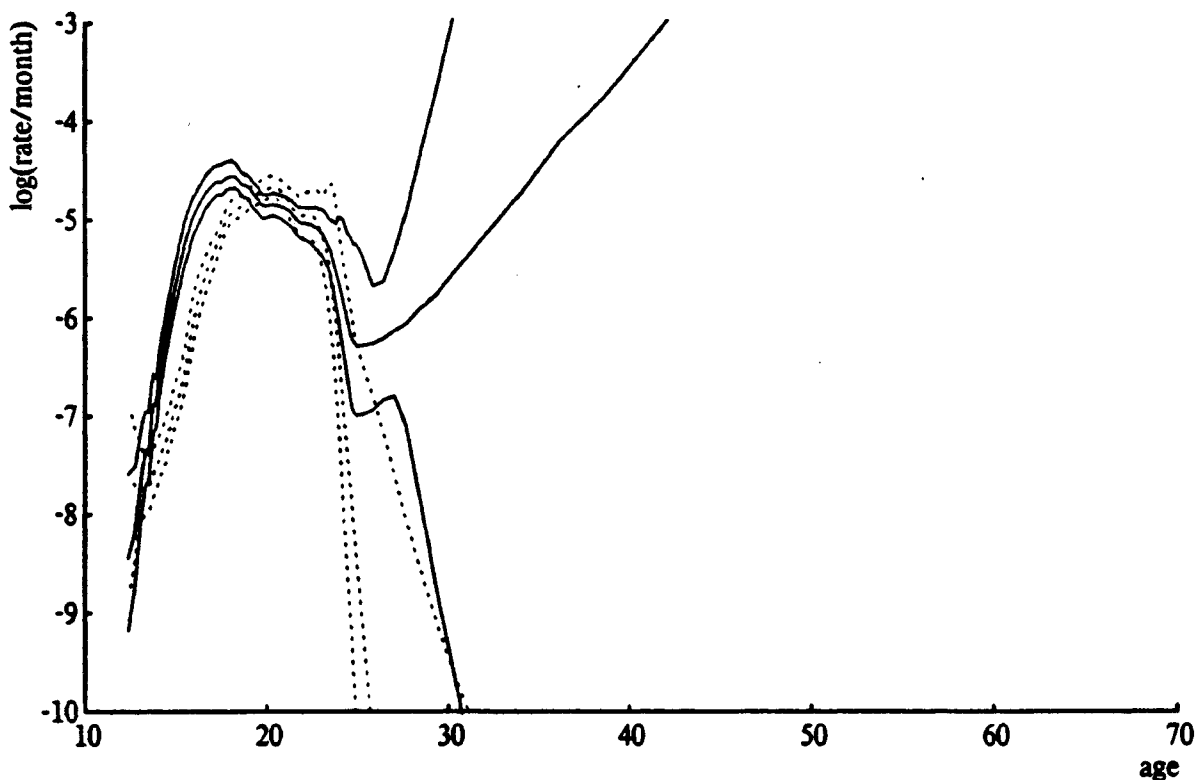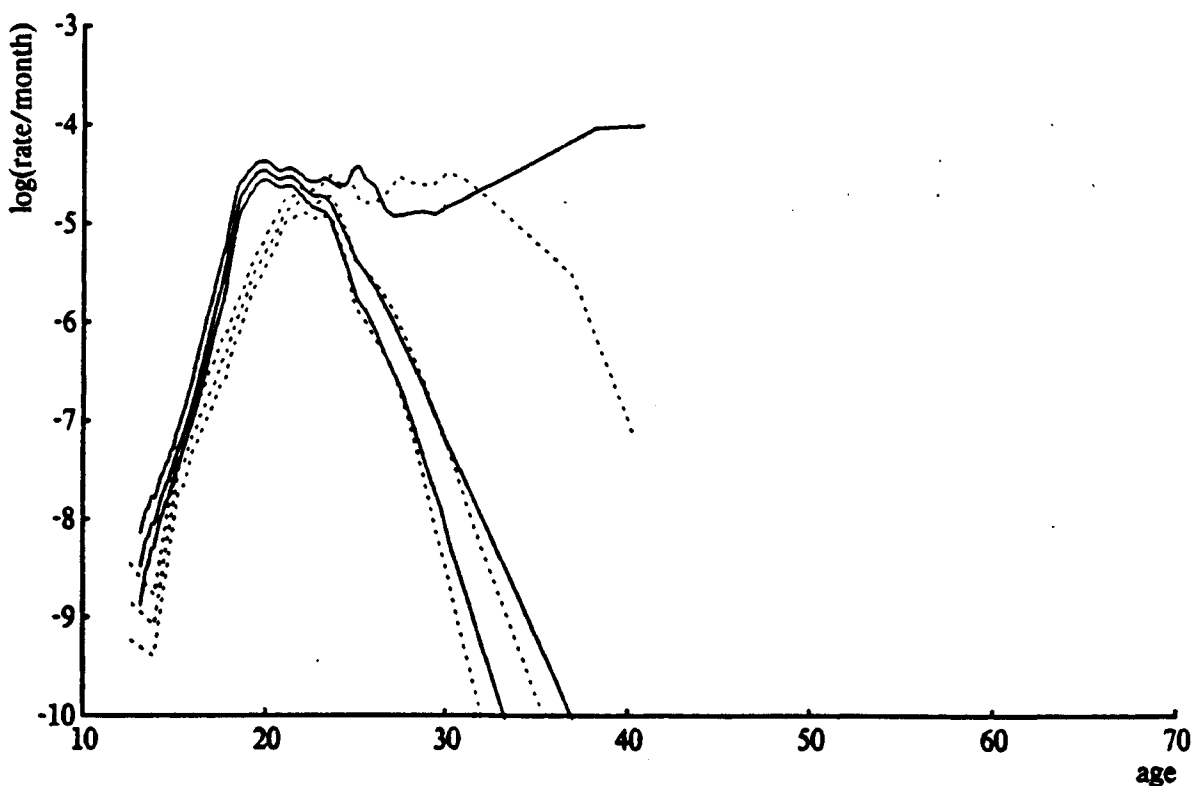
Figure 4.11a: Bootstrap Results for Smoothed Log Hazard Estimates: Completed Education for Whites (June 1980 CPS). Three curves for each group: Bias-corrected smoothed log hazard, lower, and upper bootstrap 90% pointwise confidence intervals. Solid curves for women with 0-9 years of education (I = 5976), and dotted curves for women with 10-12 years of education (I = 9072).

Figure 4.11b: Bootstrap Results for Smoothed Log Hazard Estimates: Completed Education for Whites (June 1980 CPS). Three curves for each group: Bias-corrected smoothed log hazard, lower, and upper bootstrap 90% pointwise confidence intervals. Solid curves for women with 13 years of education (I = 20578) and dotted curves for women with 14 or more years of education (I = 14513).

Figure 4.12a: Bootstrap Results for Smoothed Log Hazard Estimates: Completed Education for Blacks (June 1980 CPS). Three curves for each group: Bias-corrected smoothed log hazard, lower, and upper bootstrap 90% pointwise confidence intervals. Solid curves for women with 0–9 years of education ($I = 1246$) and dotted curves for women with 10–12 years of education ($I = 1640$).

Figure 4.12b: Bootstrap Results for Smoothed Log Hazard Estimates: Completed Education for Blacks (June 1980 CPS). Three curves for each group: Bias-corrected smoothed log hazard, lower, and upper bootstrap 90% pointwise confidence intervals. Solid curves for women with 13 years of education ($I = 1864$) and dotted curves for women with 14 or more years of education ($I = 1178$).

-4.722 at age 23.4 with a 95% confidence interval of (-4.97,-4.49) for women with 14+ years of education.

Because the log hazard estimates for blacks are subject to substantial variability and bias it is difficult to make any further black/white comparisons for the effects of education on rates of marriage. I henceforth restrict attention to the log hazard estimates for whites.

Some striking features of the log hazard estimates in Figures 4.11a and 4.11b are the observed differences in how rapidly the slope of the log hazard changes at its maximum. The log hazard curves for women with 0-9 and 13 years of education have a relative rounded appearance near the maximum in contrast to the log hazard curves for women with 10-12 and 14 years of education, which change slope near the maximum in a relatively abrupt manner. While it is relatively difficult to quantify this behavior based on the estimates in Figure 4.11, the observed differences are clearly striking.

4.4 Discussion

This section presents a discussion of three well established theoretical perspectives that inform research on marriage and the family and contrasts the empirical results of Section 4.3 with hypotheses drawn from these approaches.

4.4.1 Strength of Family Norms: Probability of Eventual Marriage

A common sociological and demographic proposition is that the mean age at marriage is lower, and the proportion ultimately marrying is higher, when individuals have a strong commitment, or are strongly socialized, to norms governing the family and family institutions (see, e.g., Hajnal, 1953). But despite widespread usage (Elder and Rockwell, 1976; Modell, Furstenberg, and Hershberg, 1976; Dixon, 1971, 1978; Gaskin, 1978; Anderson, 1984; Schoen, Baj, and Woodrow, 1984; Weir, 1984) there are great difficulties inherent in this proposition as applied to the mean age at marriage, proportion ultimately marrying, and other univariate measures for summarizing of age dependence in marital patterns. That is, univariate measures, and in particular the mean or the singulate mean age at marriage, necessarily capture only the most rudimentary aspects of how the probability of marriage varies with age, which thus limits the potential usefulness of these measure as constructs from which to base theoretical propositions.

Propositions based on the proportion ultimately marrying are, at first glance, considerably more attractive and indeed the empirical evidence tends to support the intuition that populations with higher proportion marrying are also populations that give evidence of stronger and more elaborate social mechanisms for supporting marital institutions. For example, because of the stance of the Catholic Church on matters concerning the family and marriage, researchers often hypothesize that individuals who belong to the Catholic faith have, on average, a higher individual commitment to family norms, and hence that, all other things being equal, Catholics are likely to have higher probabilities of eventual marriage than non-Catholics. The comparison of survivor curves in Figure 4.1, which shows that women of Mexican descent have the highest proportion of women married for all ages after 30, provides evidence in support of this propo-

sition under the assumption that the proportion of Catholics is lower for the white and black populations and highest for women of Mexican descent.

Nevertheless, the following argument suggests that the proportion ultimately marrying does not capture many aspects of age dependence in marital patterns that are apparent when examining the hazard rate and log hazard rate of marriage. Consider two groups, A and B, and suppose that group A has a higher proportion of individuals who ultimately marry (or who marry by age t). Then the following two hypothetical situations help demonstrate the ambiguity in focusing only on the proportion who ultimately marry when studying age dependence in rates of first marriage. Case 1: suppose that group A has a log hazard rate that is everywhere higher than group B; then it follows that group A has a higher proportion of individuals who ultimately marry (and, indeed, a higher proportion of married women at all ages.) Case 2: suppose that group A has a log hazard rate that attains a greater maximum value than group B, but that the log hazard rate for group B is greater than that for A for all ages greater than t; that is, the log hazard rates cross for some age t_1 Then as age $t \rightarrow \infty$, group B will have a higher proportion who ultimately marry; but if group A has a log hazard that attains a maximum value that is sufficiently larger than that for group B, then group A can have higher proportions who marry than group B for some range of values t. This suggests that only in Case 1 can one establish an unambiguous correspondence between the probability of marriage and statements about the proportional difference of the hazard rate if substantial age dependence exists in rates of marriage.

The estimates for log hazards in Section 4.3 suggest that only rarely do comparisons of log hazards fit Case 1. For example, the comparisons between whites and blacks, and whites and women of Mexican descent fit Case 2, where in both instances whites fit the description for group "A". However these empirical data help demonstrate the ambiguities that arise in attempting to infer proportions who ultimately marry from qualitative features of the log hazard estimates. For example, Figure 4.3 suggests that, in comparisons between whites and blacks, whites have a higher proportion who marry for all observed ages greater than 20 even though the log rate of marriage for blacks is higher in absolute terms than whites past age 35. This suggests that the asymptotic argument in Case 2 may tell us little about the proportions who marry at biologically attainable ages. On the other hand, comparisons between whites and Mexicans in Figure 4.3 show that Mexicans have a higher proportion who marry for all observed ages greater than 30, suggesting that the asymptotic conclusions reached in Case 2 *can* be relevant for biologically attainable ages.

These considerations suggest some refinements of the traditional propositions that hypothesize relations between individual attributes or group norms and the proportion of individuals who eventually marry. For example, the data for whites and women of Mexican descent in Figure 4.3 suggest that high marrying proportions are empirically consistent with at least two conceptually distinct components of the log rate: a sufficiently large maximum value for the log rate of marriage, or a sufficiently slow decline in the rate of marriage after the peak age of

marriage. Looking more closely at these data, note that the high maximum for the log rate of marriage for white women in Figure 4.3 implies that many of these women marry between the ages of 20 and 25; but because the log rate declines rapidly after 25, women who do not marry early are increasingly unlikely to marry as they grow older.[11] In contrast the significantly lower maximum for the log rate of marriage for women of Mexican descent implies that relatively fewer women marry at early ages; but because the log rate declines extremely slowly after 25, women of Mexican descent who do not marry early are much more likely to marry eventually even if they have not married early in life. Thus, if we consider a high maximum and a slow decline in the log hazard as two rough indicators for strong family norms in a subculture, then the indicators for whites and blacks are mixed and only the log hazard for women of Mexican descent provides consistent evidence for strong family norms.

The estimated log hazard rates for whites and blacks provide some clues that shed new light on the empirical differences in black/white marital patterns. The log hazard for whites has the largest maximum value for the three race-ethnic groups while the log hazard for blacks has the smallest observed maximum value; moreover the informal tests in Section 4.3.1.6 suggested that the observed differences are significant under the assumption that one can reasonably compare maximum values that occur at different ages. By this criterion we might conclude that family norms among whites were significantly stronger than those for blacks. But we also observed that the log hazard for blacks declined more slowly than the log hazard for whites for ages greater than 30; the informal tests in Section 4.3.1.6 suggested that these observed differences are also significant under the assumption that one can reasonably summarize the decline by a the slope of a linear function. Then by this criterion we might conclude that family norms among blacks were significantly stronger than those for whites.

These empirical observations suggest that cultural norms may influence marital rates in at least two distinct ways. First, cultural norms may provide a definition for the "ideal" age of marriage (Ryder and Westoff, 1971; Glick and Norton, 1977; Modell, 1980; Veroff, Douvan, and Kulka, 1981). Then I hypothesize that the corresponding log hazard will exhibit a sharper rise at ages less than the ideal age, and attain a larger maximum value at the ideal age, in direct proportion to the strength of norms concerning such an ideal age. Second, cultural norms may encourage eventual marriage. Then I hypothesize that the corresponding log hazard will exhibit a slower decline for ages past the peak ages of marriage in direct proportion to the strength of norms encouraging eventual marriage.

These interpretations suggest three quite distinct normative patterns for the race-ethnic groups examined in Figures 4.3-4.8. White women exhibit marital patterns that appear consistent with a normative subculture that provides a strong definition for an ideal age of marriage but a relatively weaker directive to marry for those women who remain unmarried after the

---

[11] These conclusions are also consistent with the ever married proportions observed in Table 4.1 for earlier white birth cohorts, in which 4-6 percent of women never marry.

ideal age of marriage. Black women exhibit marital patterns that appear consistent with a normative subculture that provides a far less strong definition for an ideal age of marriage but relatively stronger directives to marry for those women who are older and who have not yet married. Women of Mexican descent exhibit marital patterns that appear consistent with a normative subculture that provides both a strong definition for an ideal age of marriage and strong directives to marry for those women who have not yet done so.[12]

4.4.2 Strength of Family Norms: Black/White Differences

A popular explanation for low observed black rates of marriage refers to the disorganization of the black family or "defective" black commitment to family norms found in the majority culture resulting from a legacy from slavery (see, e.g., Frazier, 1939; Moynihan, 1967).[13] Against these explanations and without disputing the low absolute rates of marriage for blacks, several scholars have contested elements of the disorganization thesis (for historical studies, see, e.g., Gutman, 1976; Fogel and Engerman, 1974; Genovese, 1974; Hershberg, 1975; Engerman, 1977; for studies of the black family in the Caribbean, see, e.g., Goode, 1960, 1961; Patterson, 1982). These more recent studies have suggested that the legacy of slavery was less pronounced than previously supposed, and that differences between both the norms and observable household patterns for white and blacks in the post-emancipation period were less marked than hypothesized. This raises the possibility that other socioeconomic factors may provide more powerful explanations for observed black/white differences.

The results presented in Figures 4.3-4.8 add substance to the caveats raised about the "disorganization" thesis. On the one hand, the estimated hazard and log hazard rates provide considerable evidence that blacks have significantly lower overall rates of marriage. Similarly, the estimated survivor probabilities reported in Table 4.2 and Figure 4.1 suggest somewhat lower probabilities of eventual marriage blacks; white and black probabilities of marriage by age 45 were 86 and 87% for the 1880-1909 cohort, 95 and 90% for the 1910-1929 cohort, and 95 and 87% for the 1930-1949 cohort, respectively.[14] However, both the original smoothed estimates in Figure 4.3 and the bias corrected bootstrap results in Figure 4.8 suggested that the log rate of marriage declined less rapidly for blacks and more rapidly for whites; the conservative tests described in Section 4.3.1.7 suggested that we could reject the null hypotheses that these slopes are equal. These findings are thus consistent with the position taken by critics of the disorganization thesis that, despite circumstances that prevent more blacks from marrying,

---

[12] The conclusions for women of Mexican descent drawn from the estimates of Figure 4.3 must be qualified on the basis of the bootstrap estimates given in Figure 4.8, which suggested substantial bias and variability for log hazard estimates for the Mexican curves after age 25. On the other hand, the original smoothed estimates for women of Mexican descent appear consistent with the high marriage proportions observed in Figure 4.1, which provides some basis for the speculations I have advanced.

[13] Frazier in fact carefully distinguished between the urban and rural black family, arguing that the rural black family was more traditional in its commitment to family norms.

[14] The slightly higher probabilities of eventual marriage for blacks for the 1880-1909 cohort are not significant based on a two-sided 95% test derived from the survivor probability pointwise confidence intervals.

blacks retain a relatively strong normative commitment to marriage and marital institutions.

Some circumstantial and additional support for the last hypothesis can also be gleaned from black/white birth cohort comparisons. In Section 4.3.2 I reported preliminary evidence suggesting that, to a first order approximation, observed black/white differences in the maximum value attained by the log hazard for the 8 black and white birth cohorts could be explained by a period effect and a black/white effect for three of the four cohorts; on average, blacks had a maximum rate of marriage that was $\exp(-.75) \approx 1/2$ as large as as the maximum rate of marriage for whites. The cohort that was an exception to this empirical finding consisted of women born between 1910 and 1929, the majority of whom reached adulthood during the Depression era. I found that, for this cohort of women, the difference in the maximum values between black and white rates of marriage was noticeably smaller than that found for the other cohorts. While weak, the evidence is suggestive: if impoverished circumstances act to depress the maximum attainable value for rates of marriage, then it follows that the maximum for the black rate of marriage should more closely approach the maximum for the white rate of marriage during historical periods in which white circumstances were relatively more similar to black circumstances as compared to other historical periods.[15]

4.4.2 Human Capital

Considerable excitement has attended so-called human capital theories of the family, both from proponents (Becker, 1973; Frieden, 1973; Becker, Landes, and Michael, 1977; Becker, 1981) and critics (see, e.g., Goode, 1973; Ryder, 1973; Arthur, 1982). However there are two potential difficulties in any attempt to directly apply conventional human capital arguments to rates of first marriage. First, as Michael and Tuma note, the theoretical framework adopted by Becker is primarily a static one, and thus does not directly predict how the timing or probability of marriage might change over the life cycle.[16] Second, a counterfactual but interesting feature of the static market framework adopted by Becker is that, short of externalities or a massively long-term disequilibrium in the marketplace, the market marriage should "clear" leaving no individual unmarried within the life expectancy of the average individual.[17]

Nevertheless, there are a number of attractive aspects of the human capital theory as applied to marriage. For example, aspects of the theory have received a certain degree of

---

[15] See Engerman (1977) for a discussion of historical differences in black/white fertility and for some speculations on the effects of the Depression era on black/white fertility differentials.

[16] But see Michael and Tuma for hypotheses derived from human capital arguments that apply directly to predictions about an individual's rate of marriage at early ages.

[17] Thus, I believe that a human capital model of the marriage market would not ordinarily hypothesize effects of a marriage "squeeze" unless there was a permanent disequilibrium in the relative ages of potential mates, or if individual preferences were highly inelastic with respect to the relative age of a mate. Similarly, although the proportion of females married is typically taken to be positively related to the relative wage rate of males to females (see, e.g., Frieden, 1974), it would seem logically consistent, given the symmetry of the home production function, to suppose that females with high incomes would maximum their productivity by seeking mates with lower incomes, resulting in market production by the female and home production by the male. These arguments tend to suggest that "externalities" in marriage markets might well be sizable.

empirical support (see, e.g., Becker, Landes, and Michael, 1977; Frieden, 1973; Becker, 1981; and especially Michael and Tuma, 1986). Furthermore, the theoretical framework provided by human capital theory generates a large number of testable hypotheses (see, e.g., Becker 1981) and so has proven to be a most fertile theoretical perspective.

As an example of this theoretical fertility, it is interesting to note that the human capital perspective offers a theoretical explanation that, in retrospect, would predict a linear declining log hazard rate of marriage for ages past the peak marrying years. Recall that Becker argues that the decision to marry is primarily based upon calculations of net productivity for a prospective couple; an individual decides to marry if marginal gains are positive and remains single if marginal gains are negative. If a woman happens to remain unmarried after the peak marrying years, we can expect that she might invest relatively more heavily in alternatives to marriage such as a career. If some of the investments in a career cannot be transferred to the productivity of a couple (or if marriage may harm a woman's productivity by increasing the probability of work interruption for childbearing or decreasing the possibility of career mobility) and if the value of the "sunk" investments accumulates value with time, then the value of these investments can be expected to increase with age for every year a woman remains unmarried after the peak marrying years. Then if the rate of marriage varies negatively with the value of the nontransferable investments and assuming that the investments in career begin during the peak marrying years, we would expect the rate of marriage to decline exponentially (and the log rate of marriage to decline linearly) in direct proportion to the increasing value of the nontransferable investment.

Unfortunately, specific aspects of the argument given above do not appear entirely consistent with the empirical results of Section 4.3. Because the decision to marry is governed, in large part, by calculations based on the marginal productivity of the prospective couple, one would expect that the linear decline in the logarithm of hazard rates of marriage would respond to the relative wages of men and women (see Frieden, 1974 for an explicit statement of this assumption, with the caveat noted in footnote 16). This in turn suggests that the relative male/female wage rate is inversely proportional to a woman's log rate of marriage for ages past the prime marrying years.

Although I have not gathered time series data for the relative wage rate, it is possible to engage in some speculations about variations in the relative wage rate by ethnicity. We might suppose that the relative wage rate for Mexicans is higher than that for whites, in part based on an assumption that Mexican women have a preference for eventual marriage and against permanent market work based on Catholic norms. Then assuming low rates of interracial marriage, the comparative data for whites and Mexicans appears consistent with the human capital hypothesis: if Mexican women have a lower propensity to invest in market work then they should have a correspondingly greater inclination for investments in marriage if they have not yet married.

However one might expect that the relative wage rate for blacks is lower than that for

whites, in part based on the lower wages for black males and the higher workforce participation of black females. Then assuming low rates of interracial marriage, the human capital hypothesis would predict a more sharply declining rate of marriage for black women when compared to white women, which is contradicted by the empirical evidence and tests presented in Section 4.3. There are a number of possible explanations for these findings: 1) it may be that the predicted higher decline in the rate of marriage for blacks is offset by factors not controlled for in the analyses of the log hazards by race-ethnicity; 2) blacks may have a stronger preference for eventual marriage than do whites; 3) the a priori assumption that the relative male/female wage rates for blacks is lower than that for whites may prove incorrect if actual time series were examined; 4) the relative wage effect may be sensitive to small changes in rates of interracial marriage; or 5) relative wages may be a less powerful explanatory factor in marital decisions than is supposed by human capital theorists.

4.4.3.1 *Effects of Education on Rates of Marriage*

In Figures 4.11–4.12, I noted observed two possible regularities in the effects of education on the empirical rates of marriage. Section 4.4.3.1 discusses some common hypotheses that are consistent with the observed shifts in the age at which the rate of marriage attains its maximum. Section 4.4.3.2 draws parallels between the relative "peakedness" of rates of marriage and insights drawn from the literature that considers the effects of education as institutionalized force in the life cycle.

4.4.3.1 *Education: Shifts in the Peak Ages of Marriage*

In Figures 4.11 and 4.12 I noted a weak tendency for the age at which the log rate of marriage is a maximum to increase with years of completed education. Although weak, these results are intuitively plausible and are consistent both with standard hypotheses and empirical results obtained in recent research. Due to the standard nature of these hypotheses and results, and because of the difficulties in devising tests to assess the significance of the observed differences in the ages at which the log rate of marriage attains a maximum, I have chosen not to present an extended discussion and instead restrict discussion to a simple summary of results of previous research.

Full time school enrollment is often predicted to be incompatible with marriage. Using temporal data on the cohabitation of Swedish women, Hoem and Hoem (1985) find that being in school has strong negative effects on the rate of cohabitation, and that no significant relationship exists in their data for the effect of the current *level* of education on cohabitation after controlling for current schooling status. Similarly, Waite and Spitze (1981) suggest that the observed tendency for women with greater levels of education to marry later is due to the greater financial resources possessed by middle class families, who use these resources as a means to to discourage early marriage by offering teenagers attractive alternatives, most notably, a college education. Lastly, since norms concerning marriage tend to dictate that the couple be financially independent (see, e.g., Goode, 1970, 1982; Ryder and Westoff, 1971; Modell, 1980; Veroff, Dou-

van, and Kulka, 1981; Bellah et al., 1985), normative pressures may encourage women headed for college to delay marriage until after schooling is completed and financial circumstances are more attractive.

4.4.3.2 Education: Institutionalization of Life Course

In this section, I review recent work that reconceptualizes the role of education in more institutional terms and draw upon this literature to suggest an explanation for the variations with education in the relative "peakedness" observed in the log hazard rates for whites. This discussion is necessarily more speculative in nature because of the difficulty in assessing the significance of observed differences the relative "peakedness" of the log hazard estimates in the absence of a parametric model, and because the hypotheses I suggest have not been previously investigated in the literature.

A number of authors have commented on the role of education in institutionalizing the life course, most notably Meyer (1977, 1985). In a seminal paper, Meyer (1977) argued that education can be best viewed as an institution that acts, in some sense, to allocate and certify an individual's transition from childhood to an adult role. That is, adult roles are attributed to individuals not on the what they have learned in schools but rather on the basis of the years and types of education they receive (Meyer, 1977, pp. 58-59; see also similar arguments in Hogan, 1978; Marini, 1978). Thus, an employer is likely to weigh educational credentials such as a high school diploma heavily in hiring decisions, even when these educational credentials have little functional relationship to the work involved and even when the employer has little information about the course of study pursued by the student. Similarly, parents are more likely to disapprove of marriage for an 18 year old women if she is still attending high school because of the inappropriateness of marriage at such an early age, but are more likely to approve of marriage for the same woman half a year later after graduation, regardless of her personal maturity at either age.

More generally, these arguments suggest (Meyer, 1985) that education may play an important role in institutionalizing age roles, in the sense that the sequential completion of a standardized educational course (completing high school or four years of college) confers upon a woman an adult status that is undifferentiated with respect to all personal characteristics other than the socially-defined educational category within which she falls. Conversely while standardized normative rules empower schools with the power to grant adult status to individuals, these same rules, Meyer argues, may lower the prospects for "nonstudents"—individuals who interrupt their education, choose not to complete their education, or otherwise violate the normal—and normatively acceptable—sequential progression through educational institutions.

These considerations suggest that the more an individual's progression through educational institutions follows the normal and normatively approved sequence, the greater the likelihood that the transition to adult status is sharply demarcated by the date of graduation; conversely the more an individual's progression through educational institutions violates the normal and

normatively approved progression, the greater the ambiguity in the exact time of transition to an adult status.

The results reported in Section 4.3.3 provide some preliminary support for these hypotheses, although this statement must be tempered by the lack of any assessment of the significance of these results. Consistent with the hypothesis of a sharp demarcation between pre-adult and adult statuses by educational attainment, the estimated log hazard curves for white women with 10-12 or with 14+ years of schooling are characterized by a sharp and sudden change in the slope of the log hazard rate near typical ages of graduation. By contrast, the estimated log hazard curves for white women with less 0-9 or 13 years of schooling (and who thus have educational experiences that deviate more sharply from the prescribed normative sequence) have estimated log hazard curves that exhibit less sudden changes in slope in the peak marrying ages.

I wish to reemphasize the highly speculative nature of the comments in this section. First, as noted earlier, these speculations are based on indications without any assessment of the significance of observed differences between the log hazard rates for different educational levels. A second caveat to these results is the nature of the educational variable available in these data, which reports the education completed at the time of interview. Because a substantial number of women may have continued their education after marriage, this variable is not an accurate indicator of the level of education attained prior to marriage. Moreover, these results may be confounded with social class: if parental resources are important in influencing the timing of a child's marriage *and* for encouraging completion of education, then the effect of social class may be confounded with the effect of completing a standardized course of education. Lastly, these results are subject to selection bias in the sense suggested by Heckman in a labor setting; if persons who do not complete high school or college possess personal characteristics that make them "losers," these characteristics may also negatively influence individual attractiveness in the marriage market. Although it is most straightforward to assume that individuals who are less attractive would have lower rates of marriage at all ages (which is however not consistent with the weight of evidence in this chapter) it is conceivable that more a elaborated model of marriage market mechanisms might predict the relative roundedness or peakedness in the behavior of the log rate of marriage in a manner that depends upon the market attractiveness of an individual.

4.5 Conclusion

This chapter presented an extensive exploratory analysis of the patterns of age dependence for a sample of women in the United States. At the outset of the chapter I posed several substantive and methodological questions. With the results obtained from the exploratory analyses presented in Section 4.3, it is possible to advance some tentative conclusions to many of these questions.

4.5.1 Methodological Conclusions

In this chapter I made extensive use of the smoothed hazard estimator and bootstrap procedure introduced in Chapter 3. Given sufficiently large samples and computing resources, these proce-

dures appear to have considerable promise for exploring patterns of age dependence when there exist a sufficient number of events for the age range of interest. But for age ranges in which the number of events is sparse, the estimator may be subject to considerable variability and bias, making interpretation of the original smoothed hazard estimate difficult.

An important component of the exploratory analyses was the flexibility provided by the approximate bootstrap confidence intervals, and various informal tests proved extremely useful when attempting to roughly gauge the significance of observed group differences. Moreover, although it is true that the bootstrap method I employed required considerable computational resources, it was nevertheless possible to achieve considerable computational savings by exploiting the discrete nature of the hazard estimator. In light of the substantive gains obtained from the availability of confidence intervals, the additional computational cost of bootstrapping seems worthwhile. More generally, the results are encouraging enough to suggest that the estimator and bootstrap procedure merit further development as a useful investigative tool when the analyst wishes to make few parametric assumptions about the underlying hazard rate.

The exploratory results also presented substantial evidence suggesting that rates of marriage have a number of discernible empirical regularities. I found that these regularities are most transparently observed on a log scale, which provides clues for possible log linear parametric specifications for rates of first marriage that incorporate age dependence. In general, the qualitative shape for the log hazard of first marriage increased in a more or less linear fashion to a maximum, and decreased after this maximum in a more or less linear manner. Moreover, the qualitative shape of the log rate varied 1) in the maximum value attained by the rate; 2) in the value of the linear slopes for the ascending and descending portions of the log rate; 3) in the age at which the maximum value for the rate occurred, and 4) in the relative roundedness or peakedness (more generally, the behavior of the second derivative) of the log rate near its maximum.

These qualitative patterns provide some of the best clues to date about the manner in which the hazard rate of marriage varies with observed individual attributes, and would seem to provide a solid empirical foundation from which to develop parametric models for the rate of first marriage. For example, the overwhelming weight of evidence suggests that proportional hazard models are not useful in the context of first marriage; the log hazard estimates suggest that violations of proportionality occur with great frequency—indeed, it was rare to find any log hazard estimates that differed by a constant but common to find log hazard curves that differed in the age at which the maximum of the log rate occurred, or that differed in the slopes for the ascending or descending portions of the log rate.

4.5.2 Substantive Conclusions

While the substantive conclusions of this chapter are necessarily more speculative in nature, I feel that the exploratory results have nevertheless provided results that are extremely suggestive about the social processes governing marriage. The exploratory results provided considerable

evidence that the log hazard rate for first marriage both has a observable regular qualitative shape *and* that the observed qualitative aspects of the hazard vary in apparently quantifiable ways with observed individual attributes. These results provide overall assurance that the process of marriage is neither entirely random nor entirely deterministic in nature.

The exploratory results for race-ethnicity provided new insights that add to our understanding of black, white, and Mexican marital patterns. I reported evidence for race-ethnic differences in the maximum rate of marriage, in the speed with which the rate declines from the maximum rate, and, to a lesser extent, in the age at which the rate attains its maximum value. These results, in turn, suggest that usual hypotheses found in the literature need additional refinement if they are to account for the observed patterns found in Section 4.3 and I advanced some possible refinements for hypotheses derived from the literature on the effects of cultural norms on family institutions.

I also presented some preliminary evidence on the historical stability of black/white differences, with the noticeable exception of the Depression era cohorts, which had an observed *smaller* black/white differential. These results provide some weak evidence contradicting the popular view concerning the "defective" nature of the black family, and instead suggest that black women may have a slightly stronger predisposition toward eventual marriage relative than white women, a proposition that is consistent with more recent sociological and historical explanations of the black family.

Lastly, I presented exploratory results on the effects of education on age dependence in marital rates and provided some highly speculative comments on their possible interpretation using some insights from the recent literature on the effects of education as an institution. This literature suggests that education may serve to structure the entry into adult statuses. Some very preliminary evidence suggests that individuals who follow standard normative schooling sequences may indeed have more concentrated marital patterns. However, these findings are also possibly consistent with predictions that can be derived from theories that make no reference to institutional factors, and it is not possible to disentangle noninstitutional and institutional effects for these analyses.

References

Aalen, Odd O.
  1978 "Nonparametric Inference for a Family of Counting Processes." *Annals of Statistics* 6(4):701–726.

Anderson, J. A. and A. Senthilselvan
  1980 "Smooth Estimates for the Hazard Function." *Journal of the Royal Statistical Society* B42(3):322–327.

Anderson, Michael
  1976 "Marriage Patterns in Victorian Britain: An Analysis Based on Registration District Data for England and Wales 1861." *Journal of Family History* 1(1):55–78.
  1984 "The Social Position of Spinsters in Mid-Victorian Britain." *Journal of Family History* 9(4):377–393.

Arthur, W. Brian
  1982 "Review of Gary S. Becker, *A Treatise on the Family.*" *Population and Development Review* 8(2):393–397.

Becker, Gary S.
  1973 "A Theory of Marriage." Pp. 299–344 in Theodore W. Schultz (Ed.), *Economics of the Family: Marriage, Children, and Human Capital.* Chicago: University of Chicago Press.
  1981 *A Treatise on the Family.* Cambridge, MA: Harvard University Press.

Becker, Gary S., E. M. Landes, and Robert T. Michael
  1977 "An Economic Analysis of Marital Instability." *Journal of Political Economy* 85:1141–1187.

Bellah, Robert N., Richard Madsen, William M. Sullivan, Ann Swidler, and Steven M. Tipton
  1985 *Habits of the Heart: Individualism and Commitment in American Life.* Berkeley: University of California Press.

Borgan, Ørnulf, and Henrik Ramlau-Hansen
  1985 "Demographic Incidence Rates and Estimation of Intensities with Incomplete Information." *Annals of Statistics* 13(2):564–582.

Carr, Nancy M., Lauri Steel, and Josefina J. Card
  1983 "June 1980 U.S. Current Population Survey: Selected Variables—Females." Palo Alto: Data Archive on Adolescent Pregnancy and Pregnancy Prevention, American Institutes for Research.

Coale, Ansley J.
  1981

Coale, Ansley J., and Donald R. McNeil
  1972 "The Distribution by Age of the Frequency of First Marriage." *Journal of the American Statistical Association* 67(340):743–749.

Cox, D. R., and D. Oakes

1984 *Analysis of Survival Data.* London: Chapman and Hall.

Davis, Nancy J., and Larry L. Bumpass

1976 "The Continuation of Education After Marriage Among Women in the United States: 1970." *Demography* 13(2):161–174.

Dixon, Ruth B.

1971 "Explaining Cross-Cultural Variations in Age at Marriage and Proportions Never Marrying." *Population Studies* 25(2):215–233.

1978 "Late Marriage and Non-Marriage as Demographic Responses: Are They Similar?" *Population Studies* 32(3):449–466.

Efron, Bradley

1979 "Bootstrap Methods: Another Look at the Jackknife." *Annals of Statistics* 7(1):1–26.

1981a "Censored Data and the Bootstrap. *Journal of the American Statistical Association* 76(374):139–172.

1981b "Nonparametric Standard Errors and Confidence Intervals." *Canadian Journal of Statistics* 9(2):139–172.

1982 *The Jackknife, the Bootstrap and Other Resampling Plans.* CBMS Regional Conference Series in Applied Mathematics, 38. Philadelphia: Society for Industrial and Applied Mathematics.

1986 "Bootstrap Meets Supersmoother." Dept. of Statistics, Stanford University.

Elder, Glen H. Jr., and Richard Rockwell

1976 "The Timing of Marriage and Women's Life Patterns." *Journal of Family History* 1(1):34–53.

Engerman, Stanley L.

1977 "Black Fertility and Family Structure in the U.S., 1880–1940." *Journal of Family History* 2(2):117–138.

Finnäs, Fjalar

1980 "A Method to Estimate Demographic Intensities via Cumulative Incidence Rates." *Theoretical Population Biology* 17(3):365–379.

Fogel, Robert William, and Stanley L. Engerman

1974 *Time on the Cross: The Economics of American Negro Slavery.* Boston: Little, Brown.

Frazier, E. Franklin

1939 *The Negro Family in the United States.* Chicago: University of Chicago Press.

Freedman, David, and Persi Diaconis

1981 "On the Histogram as a Density Estimator: L_2 Theory." *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 57:453–476.

Friedan, Alan

1973 "The U.S. Marriage Market." Pp. 352–371 in Theodore W. Schultz (Ed.), *Economics of the Family: Marriage, Children, and Human Capital.* Chicago: University of Chicago

Press.

Friedman, Jerome H.

1984 "A Variable Span Smoother." Technical Report No. 5, Laboratory for Computational Statistics, Department of Statistics, Stanford University.

Friedman, Jerome H. and Robert Tibshirani

1984 "The Monotone Smoothing of Scatterplots." *Technometrics* 26(3):243-250.

Gaskin, Katharine

1978 "Age at First Marriae in Europe Before 1850: A Summary of Family Reconstitution Data." *Journal of Family History* 3(1):23-33.

Genovese, Eugene D.

1974 *Roll, Jordan, Roll: The World the Slaves Made.* New York: Patheon.

Glick, Paul C., and Arthur J. Norton

1977 "Marrying, Divorcing, and Living Together in the U.S. Today." *Population Bulletin* 32.

Goode, William J.

1960 "Illegitimacy in the Caribbean Social Structure." *American Sociological Review* 25(1): 21-30.

1961 "Illegitimacy, Anomie, and Cultural Penetration." *American Sociological Review* 26(6): 910-925.

1970 *World Revolution and Family Patterns.* New York: Free Press.

1973 "Comment: The Economics of Nonmonetary Variables" (comment on "A Theory of Marriage," Gary S. Becker). Pp. 345-351 in Theodore W. Schultz (Ed.), *Economics of the Family: Marriage, Children, and Human Capital.* Chicago: University of Chicago Press.

1982 *The Family.* 2nd Edition. Englewood Cliffs, New Jersey: Prentice Hall.

Gutman, Herbert G.

1976 *The BLack Family in Slavery and Freedom, 1750-1925.* New York: Pantheon.

Hershberg, Theodore

1975 "Free-Born and Slave-Born Blacks in Antebellum Philadelphia." Pp. 395-426 in Stanley L. Engerman and Eugene D. Genovese, (Eds.), *Race and Slavery in the Western Hemisphere: Quantitative Studies.* Princeton: Princeton University Press.

Hajnal, J.

1953 "Age at Marriage and Proportions Marrying." *Population Studies* 7:111-136.

1965 "European Marriage Patterns in Perspective." Pp. 101-143 in D. V. Glass and D. E. C. Eversley (Eds.), *Population in History: Essays in Historical Demography.* London: Edward Arnold.

1983 "Two Kinds of Pre-Industrial Household Formation System." Pp. 65-104 in Richard Wall (Ed.), *Family forms in historic Europe.* Cambridge: Cambridge University Press.

Hobcroft, John, Jane Menken, and Samuel Preston

    1982   "Age, Period, and Cohort Effects in Demography: A Review." *Population Index* 48(1):4-43.

Hoem, Jan M.

    1972   "On the Statistical Theory of Analytic Graduation." Pp. 569-600 in L. M. LeCam, J. Neyman, and E. L. Scott (Eds.), *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability.*

    1976   "The Statistical Theory of Demographic Rates: A Review of Current Development (with discussion)." *Scandinavian Journal of Statistics* 3(4):169-185

Hoem, Jan, and Britta Hoem

    1985   "Life Phases of Swedish Women." Paper presented at the Meetings of the NIMH-NIA Network on Longitudinal Research Methodology. Palo Alto, California, May, 1985.

Hoem, Jan M., Dan Madsen, Jørgen Løvgreen Nielsen, Else-Marie Ohlsen, Hans Oluf Hansen, and Bo Rennermalm

    1980   "Experiments in Modelling Recent Danish Fertility Curves." *Demography* 18(2):231-244.

Hogan, Dennis P.

    1978   "The Variable Order of Events in the Life Course." *American Sociological Review* 43(4):573-586.

Johnson, Norman L., and Samuel Kotz

    1970   *Continuous Univariate Distributions-I.* New York: Wiley.

Lawless, J. F.

    1982   *Statistical Models and Methods for Lifetime Data.* New York: Wiley.

Marini, Margaret Mooney

    1978   "The Transition to Adulthood: Sex Differences in Educational Attainment and Age at Marriage." *American Sociological Review* 43(4):483-507.

Meyer, John W.

    1977   "The Effects of Education as an Institution." *American Journal of Sociology* 83(1):55-77.

    1985   "The Self and the Life Course: Institutionalization and Its Effects." In A. Sørensen, F. Weinert, and L. Sherod (Eds.), *Human Development and the Life Course.* Hillsdale, N.J.: Erlbaum.

Michael, Robert T., and Nancy Brandon Tuma

    1985   "Entry into Marriage and Parenthood by Young Men and Women: The Influence of Family Background." *Demography* 22(4):515-544.

Miller, Rupert G.

    1981   *Survival Analysis.* New York: Wiley.

Modell, John

1980 "Normative Aspects of American Marriage Timing Since World War II." *Journal of Family History* 5(2):210-234.

Modell, John, Frank Furstenberg, and Theodore Hershberg

1976 "Social Change and Transitions to Adulthood in Historical Perspective." *Journal of Family History* 1(1):7-32.

Moynihan, Daniel P.

1967 "The Negro Family: The Case fo National Action." Pp. 41-124 in Lee Rainwater and William L. Yancey (Eds.), *The Moynihan Report and the Politics of Controversy.* Cambridge, Mass: MIT Press.

Nelson, Wayne

1972 "Theory and Applications of Hazard Plotting for Censored Failure Data." *Technometrics* 14(4):945-966.

Parr, William C.

1985 "The Bootstrap: Some Large Sample Theory and Connections with Robustness." *Statistics and Probability Letters* 3(2):97-100.

Patterson, Orlando

1982 "Persistence, Continuity, and Change in the Jamaican Working-Class Family." *Journal of Family History* 7(2):135-161.

Ramlau-Hansen, Henrik

1983 "Smoothing Counting Process Intensities by Means of Kernel Functions." *Annals of Statistics* 11(2):453-466.

Rice, John, and Rosenblatt, Murray

1976 "Estimation of the Log Survivor Function and Hazard Function." *Sankhya* 38(A, Part I):60-78.

Romano, Joseph P.

1986 "On Bootstrapping the Joint Distribution of the Location and Size of the Mode." Dept. of Statistics, University of California, Berkeley, California.

Ryder, Norman B.

1965 "The Cohort as a Concept in the Study of Social Change." *American Sociological Review* 30(6):843-861.

1973 "Comment on 'Economic Theory of Fertility Behavior,' Robert J. Willis." Pp. 76-80 in Theodore W. Schultz (Ed.), *Economics of the Family: Marriage, Children, and Human Capital.* Chicago: University of Chicago Press.

Ryder, Norman, and Charles F. Westoff

1971 *Reproduction in the United States, 1965.* Princeton: Princeton University Press.

Scott, David W.

1985 "Averaged Shifted Histograms: Effective Nonparametric Density Estimators in Several Dimensions." *Annals of Statistics* 13(3):1024-1040.

Schoen, Robert, John Baj, and Karen Woodrow

    1984   "Marriage and Divorce in Twentieth-Century Belgian Cohorts." *Journal of Family History* 9(1):88–103.

Schenker, Nathaniel

    1985   "Qualms about Bootstrap Confidence Intervals." *Journal of the American Statistical Association* 80(390):360–361.

Silverman, B. W.

    1984   "A Fast and Efficient Cross-Validation Method for Smoothing Parameter Choice in Spline Regression." *Journal of the American Statistical Association* 79(387):584–589.

    1985   "Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting." *Journal of the Royal Statistical Society* B47(1):1–52.

Tanner, Martin A.

    1984   "Algorithm AS 202: Data-based Non-parametric Hazard Estimation." *Applied Statistics* 33(2):248–258.

Tanner, Martin A., and Wing Hung Wong

    1984   "Data-based Nonparametric Estimation of the Hazard Function with Applications to Model Diagnostics and Exploratory Analysis." *Journal of the American Statistical Association* 79(385):174–182.

Tuma, Nancy Brandon

    1982   "Nonparametric and Partially Parametric Approaches to Event-History Analysis." *Sociological Methodology 1982*, San Francisco: Jossey-Bass.

Tuma, Nancy Brandon, and Michael T. Hannan

    1984   *Social Dynamics: Models and Methods*. Orlando: Academic Press.

United States Bureau of the Census

    1977   *The Current Population Survey: Design and Methodology*. Technical Paper No. 40, U.S. Bureau of the Census.

    1983   *June 1980 U.S. Current Population Survey: Selected Variables—Females*. [Machine-readable data file, 71,407 logical records.] First DAAPPP 1983 edition. Palo Alto, California: Data Archive on Adolescent Pregnancy and Pregnancy Provention, American Institutes for Research.

Veroff, Joseph, Elizabeth Douvan, and Richard Kulka

    1981   *The Inner American: A Self-Portrait from 1957 to 1976*. New York: Basic.

Waite, Linda J., and Glenna D. Spitze

    1981   "Young Women's Transition to Marriage." *Demography* 18(4):681–694.

Weir, David R.

    1984   "Rather Never than Late: Celibacy and Age at Marriage in English Cohort Fertility." *Journal of Family History* 9(4):340–354.

Yandell, Brian S.

    1983   "Nonparametric Inference for Rates with Censored Survival Data." *Annals of Statistics* 11(4):1119–1135.

Education and migration

Michael Wagner

1. Problem

Der vorliegende Beitrag untersucht theoretisch wie empirisch den
Zusammenhang zwischen Bildung und Migration in dreifacher Hinsicht:
(1) In der Bundesrepublik Deutschland variiert das Bildungsniveau
der Bevölkerung regional. Der Anteil der Höherqualifizierten an der
Bevölkerung steigt mit der Ortsgröße und Siedlungsdichte (MAMMEY
1979: 152, BIRG 1985: 23). Für dieses räumliche Bildungsgefälle
gibt es zwei Erklärungen:
  a) bildungsselektive Migrationen führen zu einer Konzentration der
     Höherqualifizierten in den Verdichtungsräumen;
  b) Individuen, die in größeren Städten aufwachsen, erhalten eine
     bessere Ausbildung als Individuen, die in ländlichen Regionen
     aufwachsen.
Die These, daß die regionale Verteilung der Bevölkerung nach ihrem
Bildungsniveau durch bildungsselektive Migrationen bedingt sei, hat
in der Migrationsforschung Tradition: Mit nicht gerade wertfreien
Begriffen wie "Landflucht" (vgl. ALBRECHT 1972: 47 ff.), "Auslese-
wirkungen" (HEBERLE 1936: 170 f.) oder "Siebungsvorgängen"
(HORSTMANN 1976: 128, 133) wurde jener Tatbestand schon vor vielen
Jahren beschrieben.

Das zweite Argument beruht auf der Annahme, daß Verdichtungsräume
mit Ausbildungseinrichtungen besser ausgestattet sind als ländliche
Räume. Großstädtern aufgrund spezifischer Sozialisationsbedingungen
im Elternhaus höhere Bildungsaspirationen zu unterstellen, ver-
schiebt das Problem lediglich auf die Frage, warum das Bildungs-
niveau von Eltern in Großstädten besonders hoch ist. Die Annahme,
daß die lokale Infrastruktur im Bildungsbereich einen bedeutsamen
Einfluß auf die Bildungsbeteiligung hat, führte in der Bundesre-
publik Deutschland zu politischen Maßnahmen, die einen Abbau der
regionalen Disparitäten bei weiterführenden Bildungseinrichtungen
zum Ziel hatten. Nicht zuletzt wurde in den 60er Jahren auch dem
"Fahrschülerproblem" größere Aufmerksamkeit geschenkt (GEIPEL 1965:
14 ff.).

(2) Das Bildungsniveau hat, wie zahlreiche Studien belegen, einen
deutlichen Einfluß auf das Ausmaß räumlicher Mobilität. Für die
Bundesrepublik Deutschland belegen dieses die Arbeiten von HOFBAUER
& NAGEL (1973), SCHREIBER (1975) und MAREL (1980), in demogra-
phischen Studien aus den USA wurde dieser Nachweis erbracht (BOGUE
& HAGOOD 1953, SHYROCK & NAM 1965, SUVAL & HAMILTON 1964,
FRIEDLANDER & ROSHIER 1966, MORRISON 1972, LONG 1973), ebenso in
Großbritannien (KIERNAN 1977) und Frankreich (COURGEAU 1984, 1985)
Lediglich SANDEFUR & SCOTT (1981) kamen im Hinblick auf die Bil-
dungsselektivität räumlicher Mobilität zu einem negativen Ergebnis.
Die Autoren hatten retrospektiv erhobene Längsschnittdaten aus dem
Jahre 1969 zur Verfügung. Selbst auf Migrationsraten zwischen
Bundesstaaten der USA hatte die Bildungsvariable keinen signifikan-
ten Einfluß. Die Autoren führten diese Inkonsistenz zu den bis
dahin gewonnenen Forschungsergebnissen darauf zurück, daß Quer-
schnittdaten den Effekt von Bildung auf räumliche Mobilität
überschätzen würden. Dieses Argument überzeugt indessen nicht, zu-
mal KIERNAN (1977) den Einfluß der Bildung auf räumliche Mobilität
mit Hilfe von Längsschnittdaten belegt.
Allerdings sind die Forschungsergebnisse unter zum Teil wesentli-
chen Gesichtspunkten uneinheitlich. Erstens ist nicht sicher, ob
sich die Migrationsraten der Bildungsgruppen auch bei Nah- oder
intraregionalen Wanderungen unterscheiden (QUIGLEY & WEINBERG
1977). Zweitens gibt es kaum empirische Analysen, in denen die Bil-
dungsselektivität der Migration in Abhängigkeit von Merkmalen der

Das zweite Argument beruht auf der Annahme, daß Verdichtungsräume mit Ausbildungseinrichtungen besser ausgestattet sind als ländliche Räume. Großstädtern aufgrund spezifischer Sozialisationsbedingungen im Elternhaus höhere Bildungsaspirationen zu unterstellen, verschiebt das Problem lediglich auf die Frage, warum das Bildungsniveau von Eltern in Großstädten besonders hoch ist. Die Annahme, daß die lokale Infrastruktur im Bildungsbereich einen bedeutsamen Einfluß auf die Bildungsbeteiligung hat, führte in der Bundesrepublik Deutschland zu politischen Maßnahmen, die einen Abbau der regionalen Disparitäten bei weiterführenden Bildungseinrichtungen zum Ziel hatten. Nicht zuletzt wurde in den 60er Jahren auch dem "Fahrschülerproblem" größere Aufmerksamkeit geschenkt (GEIPEL 1965: 14 ff.).

(2) Das Bildungsniveau hat, wie zahlreiche Studien belegen, einen deutlichen Einfluß auf das Ausmaß räumlicher Mobilität. Für die Bundesrepublik Deutschland belegen dieses die Arbeiten von HOFBAUER & NAGEL (1973), SCHREIBER (1975) und MAREL (1980), in demographischen Studien aus den USA wurde dieser Nachweis erbracht (BOGUE & HAGOOD 1953, SHYROCK & NAM 1965, SUVAL & HAMILTON 1964, FRIEDLANDER & ROSHIER 1966, MORRISON 1972, LONG 1973), ebenso in Großbritannien (KIERNAN 1977) und Frankreich (COURGEAU 1984, 1985) Lediglich SANDEFUR & SCOTT (1981) kamen im Hinblick auf die Bildungsselektivität räumlicher Mobilität zu einem negativen Ergebnis. Die Autoren hatten retrospektiv erhobene Längsschnittdaten aus dem Jahre 1969 zur Verfügung. Selbst auf Migrationsraten zwischen Bundesstaaten der USA hatte die Bildungsvariable keinen signifikanten Einfluß. Die Autoren führten diese Inkonsistenz zu den bis dahin gewonnenen Forschungsergebnissen darauf zurück, daß Querschnittdaten den Effekt von Bildung auf räumliche Mobilität überschätzen würden. Dieses Argument überzeugt indessen nicht, zumal KIERNAN (1977) den Einfluß der Bildung auf räumliche Mobilität mit Hilfe von Längsschnittdaten belegt.
Allerdings sind die Forschungsergebnisse unter zum Teil wesentlichen Gesichtspunkten uneinheitlich. Erstens ist nicht sicher, ob sich die Migrationsraten der Bildungsgruppen auch bei Nah- oder intraregionalen Wanderungen unterscheiden (QUIGLEY & WEINBERG 1977). Zweitens gibt es kaum empirische Analysen, in denen die Bildungsselektivität der Migration in Abhängigkeit von Merkmalen der

Herkunfts- oder Zielregion betrachtet wird. So ist empirisch
keineswegs gesichert, daß es vor allem die Land-Stadt-Wanderungen
sind, an denen Höherqualifizierte überproportional beteiligt sind.
Eine orientierende Annahme dieser Untersuchung ist jedoch, daß der
Verdichtungsgrad von Regionen wesentlich beeinflußt, welches Bil-
dungsniveau die Personen haben, die zu- bzw. abwandern.
Die Ergebnisse der erwähnten Studien sind zudem schwerer vergleich-
bar, als es erscheinen mag. So wird die Klassifizierung der Wande-
rungen nach Distanzen uneinheitlich und in zum Teil grober Art an-
hand von Verwaltungsgrenzen vorgenommen. Ferner ist die Messung des
Bildungsniveaus unterschiedlich. In einigen Studien wird es durch
die Art des Schulabschlusses operationalisiert, andere verwenden
die Bildungsdauer oder messen Intelligenzquotienten. Letztere wer-
den von ALBRECHT (1972: 126) so resümiert: "Es kann als sicher
angesehen werden, daß eine höhere Intelligenz mit größerer
Wahrscheinlichkeit zu einer stärkeren geographischen Mobilität
führt als eine durchschnittliche oder unterdurchschnittliche
Intelligenz. Über die Richtungen dieser Wanderungen kann man sagen,
daß sie in der Regel auf die größeren Städte gerichtet sein dürften
(...)".
(3) Für eine theoretisch angeleitete Analyse der Beziehungen
zwischen dem Bildungsniveau und räumlicher Mobilität ist es
unumgänglich zu wissen, ob sich Bildungseffekte auch dann
nachweisen lassen, wenn individuelle Merkmale berücksichtigt wer-
den, die häufig mit dem Bildungsgrad eng zusammenhängen. So ermit-
telten LANSING & MUELLER (1967), daß der Bildungseffekt aufgehoben
wird, wenn der Einfluß der beruflichen Stellung statistisch kon-
trolliert wird. Offen ist zudem, ob der Einfluß des Bildungsnivaus
mit dem Alter variiert. LONG (1973) zeigte, daß die Bildungsselek-
tivität der Migration in den USA mit zunehmendem Alter abnimmt;
Bildung hätte demnach auf räumliche Mobilität nur einen kurzzei-
tigen Einfluß, zum Beispiel in den frühen Phasen des Erwerbsver-
laufs. COURGEAU (1984) belegte jedoch mit für Frankreich repräsen-
tativen Längsschnittdaten, daß bei statistischer Kontrolle des
Alters der Bildungseffekt erhalten bleibt.
Die Migrationsforschung hat die Erklärung der Bildungsselektivität
von Wanderungen bislang aus vier Perspektiven vorgenommen. Dabei
wurden entweder regionale Disparitäten im Bildungsbereich oder auf
dem Arbeitsmarkt betont (ALBRECHT 1972: 87, HOFBAUER & NAGEL 1973:
259, LANSING & MUELLER 1967: 44), ein mit dem Bildungsniveau vari-

ierendes Ausmaß lokaler Bindungen (MUSGROVE 1963), unterschiedliche
Informationsniveaus (SHAW 1975, GRUBEL & SCOTT 1967) oder Wertmu-
ster (SUVAL & HAMILTON 1964 :547). Diese Thesen sollen nachfolgend
dargestellt und diskutiert werden.

2. Theoretische Ansätze

2.1 Regionale Ungleichheiten bei Bildungseinrichtungen und auf dem Arbeitsmarkt

Regionale Disparitäten im Bildungsbereich, insbesondere dem der
Berufsausbildung, bestimmen Ausmaß und Richtung von Migrationen,
die dem Ziel dienen, eine spezifische Ausbildung zu absolvieren.
Läßt man die "klassischen", kleineren Universitätsstädte beiseite,
konzentrieren sich berufsbildende Einrichtungen - des staatlichen
und privaten Sektors - in den Verdichtungsräumen (DERENBACH 1984:
26). Entsprechende Wanderungen werden jedoch nicht zahlreich sein:
Geringe finanzielle Ressourcen der Personen, die sich in der Aus-
bildung befinden, ein niedriges Informationsniveau und die erfor-
derliche, frühe Trennung vom Elternhaus werden derartige
Wohnortwechsel eher erschweren. In der Bundesrepublik Deutschland
wurden räumliche Disparitäten bei weiterführenden Bildungseinrich-
tungen, von PEISERT (1967) detailliert dokumentiert, in den 60er
und 70er Jahren abgebaut. Bei Realschulen und Gymnasien stieg die
Zahl der Standorte zwischen 1960 und 1980 von 3.000 auf 5.100, der
durchschnittliche Einzugsbereich verringerte sich von 84 qkm auf 49
qkm (ROB 1982: 87). Allerdings ist der Einfluß des Staates auf die
Standortwahl von Unternehmen, die betriebliche Ausbildungsplätze
zur Verfügung stellen, geringer. Noch in den 80er Jahren bestehen
bei der Versorgung mit beruflichen Ausbildungsplätzen erhebliche
regionale Unterschiede (DERENBACH 1984: 3, 28; ROB 1986: 159).
Selbst dann, wenn die regionalen Ungleichheiten im Bildungsbereich
reduziert wurden, ist jedoch offen, ob dadurch ausbildungsmotivier-
te Wanderungen abgenommen haben. Denkbar wäre, daß sich dadurch
lediglich die Bildungsbeteiligung erhöht hat.
Die zweite These behauptet, daß vor allem Personen mit einem hohen
Bildungsniveau in Teilarbeitsmärkte eintreten, die regional weit
ausgedehnt sind (HOFBAUER & NAGEL 1973: 259, ALBRECHT 1972: 87,

LANSING & MUELLER 1967: 44). Je höher das Qualifikationsniveau von
beruflichen Positionen ist, desto weiter sind ihre Standorte
voneinander entfernt. Dieses ist deshalb der Fall, weil

a) die Anzahl der Arbeitsplätze für Höherqualifizierte relativ
niedrig ist,

b) diese Arbeitsplätze sich in den Verdichtungsräumen konzentrie-
ren und

c) die Verdichtungsräume über die Siedlungsfläche der Bundesrepu-
blik Deutschland relativ gleichmäßig verteilt sind.

Daß sich Arbeitsplätze der Höherqualifizierten in größeren Städten
konzentrieren, erscheint aus mehreren Gründen plausibel zu sein.
Zum einen haben Betriebe des tertiären Sektors ihren Standort vor
allem in Verdichtungsräumen (ROB 1978: 19), ein großer Anteil der
Personen mit hohem Bildungsniveau arbeitet aber in diesem
Wirtschaftsbereich (BLOSSFELD 1985b: 104 ff.). Zudem ist der Grad
der beruflichen Differenzierung in Verdichtungsräumen höher als in
ländlichen Regionen. Gerade für spezialisierte Tätigkeiten sind
Personen mit einem hohen Bildungsniveau jedoch am ehesten qualifi-
ziert. Und schließlich ist darauf hinzuweisen, daß Großstädte Zen-
tren ökonomischer und politischer Macht darstellen. Aufgrund von
Agglomerationsvorteilen und einer guten verkehrs- und kommunika-
tionstechnischen Infrastruktur, konzentrieren sich hier Unternehmen
mit statushohen Positionen, die an die Positionsinhaber hohe Quali-
fikationsanforderungen stellen.

2.2 Lokale und soziale Bindungen

In der Migrationsforschung ist vielfach nachgewiesen worden, daß
mit der Stärke lokaler Bindungen die Wahrscheinlichkeit, den
Wohnstandort zu wechseln, sinkt. Es fragt sich nun, ob Höherquali-
fizierte nur geringe lokale Bindungen haben und deshalb räumlich
mobiler sind als weniger Qualifizierte. Als Begründung hierfür wird
genannt, daß mit zunehmendem Bildungsniveau die Wahrscheinlichkeit
zunimmt, daß der Herkunftsort zu Ausbildungszwecken verlassen
wurde. In einer 1973/74 durchgeführten Untersuchung der Ellwanger
Abiturienten gaben von 1228 Befragten 83,6 % an, "gleich nach dem
Abitur weggegangen" zu sein - vor allem in die nächstgelegene
Universitätsstadt (NIEDZWETZKI 1979: 17). Hinzuzufügen wäre, daß
auch räumliche Mobilität bei Berufseintritt lokale Bindungen ver

mindert und damit weitere Wanderungen im Zuge des Erwerbsverlaufs
erleichtert.

Da Personen mit einem hohen Bildungsniveau im Sinne der Humankapi-
taltheorie (BECKER 1964) viel in ihre Ausbildung investiert haben,
müßten sie bei einem Wechsel des Berufs beträchtliche Kosten auf
sich nehmen. Bieten sich keine adäquaten Beschäftigungsmöglichkei-
ten am Wohnort, ist ein Berufswechsel daher unwahrscheinlich, eine
Migration jedoch, kann die "Erträge" der Bildungsinvestition
sichern. Eine lange und qualifizierte Ausbildung führt zu einem
beruflichen "commitment" (vgl. BECKER 1960, FORD 1973), Personen
sind beruflich "festgelegt" und müssen ihr Verhalten den regional
unterschiedlichen Beschäftigungsmöglichkeiten eher anpassen.

2.3 Informationsniveau und Wertmuster

Zwei weitere Hypothesen interpretieren die Bildungseffekte auf
räumliche Mobilität mit Hilfe von Konzepten, die handlungstheoreti-
schen Ansätzen entstammen. So meinen SHAW (1975) und GRUBEL & SCOTT
(1967), daß mit dem Bildungsgrad das Informationsniveau über alter-
native Wohnorte und -gelegenheiten zunehmen würde. Dagegen vermuten
SUVAL & HAMILTON (1964: 547), daß sich mit dem Bildungsniveau die
Aktzeptanz von Mittelschichtwerten, wie Leistungsbereitschaft oder
Pflichterfüllung, erhöht. Wanderungen, vor allem über weitere Di-
stanzen, werden in Kauf genommen, wenn es der Berufskarriere dient.
Als Vermittler von Wissen und Werten nehmen Schulen eine zentrale
Funktion ein. Daß dieses auch räumliche Konsequenzen für Individuen
haben kann, verdeutlicht MUSGROVE. Er argumentiert, daß es sogar
ein Ziel schulischer Ausbildung sei, Kindern Lebenschancen
unabhängig von ihrer sozialen und regionalen Herkunft zu verschaf-
fen: The family should be made an irrelevance: the child's life
chances should be independent of his geographical and kinship
connexions" (1963: 88). Es sind, so MUSGROVE, vor allem die
weiterführenden Schulen, die, ökonomischen Erfordernissen ent-
sprechend, Kinder auf ein Berufsleben mit Wohnortwechseln vorberei-
ten: "Today selection for a grammer-school education is selection
for a probably migratory future. The modern grammer school is an
agency for collecting local talent from its region and redeploying

it nationally and even internationally" (1963: 89). Der Einfluß des Informationsniveaus, von Werten und Einstellungen auf räumliche Mobilität ist in erster Linie Folge einer Sozialisation, die die subjektiven Bedingungen schafft, um die räumliche Allokation von Erwerbstätigen in einem ökonomischen System mit regionalen Disparitäten zu gewährleisten.

3. Datenbasis und Methoden

Die hier verwandten Daten wurden im Rahmen des Projektes "Lebensverläufe und Wohlfahrtsentwicklung" im Sonderforschungsbereich 3 der Deutschen Forschungsgemeinschaft zwischen Oktober 1981 und Februar 1983 erhoben. Die Population der retrospektiven Repräsentativerhebung bezieht sich auf Frauen und Männer der Geburtsjahrgänge 1929-1931, 1939-1941 und 1949-1951 mit deutscher Staatsangehörigkeit in der Bundesrepublik und West-Berlin in privaten Haushalten. Die Stichprobe wurde auf der Grundlage einer Erfassung von 13.974 Haushaltslisten in 404 Stimmbezirken gezogen. Sie war geschichtet nach Geschlecht und Geburtskohorte. Insgesamt liegen retrospektive Lebensläufe für 2171 Befragte vor. Die Feldarbeit wurde von GETAS (Bremen) unter Betreuung des Zentrums für Umfragen, Methoden und Analysen (ZUMA, Mannheim) durchgeführt. Zur Methodik der Untersuchung liegen mehrere Berichte vor (BRÜCKNER et al. 1984, PAPASTEFANOU 1980, TÖLKE 1982, WIEDENBECK 1981). Hervorzuheben ist, daß auch die räumliche Repräsentativität hinsichtlich der Verteilung nach Bundesländern und Ortsgrößenklassen sich bei Vergleichen mit Mikrozensusdaten als außerordentlich gut herausgestellt hat (BLOSSFELD 1985a).
Die Befragung richtete sich auf die retrospektive Erhebung zentraler Lebensbereiche entlang einer kontinuierlichen Zeitachse. Sie bezog sich schwergewichtig auf die Familien- und Wohngeschichte sowie den Ausbildungs- und Berufsverlauf.
Als Migration wird jeder von den Befragten angegebene Wohnungswechsel bezeichnet. Für die Unterscheidung von Nah- und Fernwanderungen wurde ein Schwellenwert von 50 km gewählt. Der Wahl dieses Wertes liegt die Annahme zugrunde, daß Wohnungswechsel, die unter einer Distanz von 50 km liegen, ein tägliches Pendeln zum Arbeitsplatz (gerade) noch zulassen. Informationen über die Größe von Or-

ten beruhen auf Angaben der Befragten. Wohnorte konnten als Weiler,
Dorf, Klein-, Mittel- oder Großstadt eingestuft werden. Zur Messung
des Bildungsniveaus wird der höchste, allgemeinbildende Schulab-
schluß verwendet. Unterschieden werden: a) höchstens Volksschulab-
schluß, b) Mittlere Reife, c) Fachhochschulreife oder Abitur.
Multivariate Analysen der Bildungsselektivität von Migrationen wer-
den mit Hilfe dynamischer Ratenmodelle durchgeführt. In diesen
Modellen ist die abhängige Variable die Rate (Hazard- oder Über-
gangsrate) eines Wohnungswechsels. Sie wird u.a. in Abhängigkeit
von individuellen Merkmalen (Bildungsniveau, Geschlecht u.a.) spe-
zifiziert und von besonderem Interesse ist, welche Faktoren die
Geschwindigkeit oder die Richtung (Land/Stadt) von Wohnungs-
wechseln beeinflussen. Analyseeinheiten sind Wohnperioden. Enden
sie nicht mit einem Wohnungswechsel (sondern mit dem Interview),
handelt es sich um Zensierungen. Ich verwende hier das Proportio-
nal-Hazards-Regressionsmodell von COX (1972, 1975):

$$r(t/\underline{x}) = b(t) \exp(\underline{x}'\underline{\beta}).$$

In diesem Modell ist $r(t/\underline{x})$ die Übergangsrate, $b(t)$ eine Basis-
Übergangsrate, die nicht spezifiziert werden muß, von der aber
angenommen wird, daß sie für alle Individuen gleich ist, $\underline{x}'$ ist ein
Kovariatenvektor, mit $\underline{\beta}$ werden die jeweiligen Regressionskoeffi-
zienten und mit t die Wohndauer bezeichnet. Die mathematischen
Grundlagen dieser statistischen Verfahren werden in den Arbeiten
von TUMA & HANNAN (1984) und BLOSSFELD et al. (1986) ausführlich
beschrieben.

4. Empirische Ergebnisse

4.1 Bildungsselektive Land-Stadt-Wanderungen oder Infrastruktur-
effekt?

Ich möchte nun die bereits erwähnten Thesen aufgreifen, die sich
auf den höheren Anteil von besser ausgebildeten Personen in den
größeren Städten beziehen. Zum einen wurde behauptet, diese regio-
nal ungleiche Verteilung sei eine Folge bildungsselektiver Land-
Stadt-Wanderungen, die Gegenbehauptung verweist auf die bessere
Ausstattung der Verdichtungsräume mit Bildungseinrichtungen. Die

Abbildungen 1 und 2 zeigen für zwei Bildungsgruppen und drei
Geburtsjahrgänge, wie groß der altersspezifische Anteil von Perso-
nen ist, der in Mittel- oder Großstädten lebt.
Ein erster Befund betrifft die regionale Herkunft von Personen un-
terschiedlicher Bildungsgruppen. Man sieht sehr deutlich, daß
diejenigen mit einem höheren Bildungsniveau eher in größeren
Städten aufgewachsen sind, als es bei Personen mit einem niedri-
geren Bildungsniveau der Fall ist. Außerdem nimmt der Anteil der
Personen mit höheren Bildungsabschlüssem, die in Mittel- oder
Großstädten wohnen, bis etwa zum 25. Lebensjahr viel kräftiger zu
als bei Absolventen der Volksschule. Beide der eingangs genannten
Thesen werden also unterstützt: Man beobachtet bis zur Mitte des
dritten Lebensjahrzehnts eine bildungsselektive Land-Stadt-Wande-
rung, andererseits stammen die Höherqualifizierten in einem hohem
Ausmaß aus Regionen mit relativ guten Bildungsangeboten.
Kohortenunterschiede zeigen sich vor allem bei den besser Qualifi-
zierten in den ersten zwei Lebensjahrzehnten. Insbesondere die
Auswirkungen des Zweiten Weltkrieges sind bei den um 1930 und 1940
Geborenen unübersehbar. Es ist bemerkenswert, daß Personen mit
einer (späteren) höheren Schulausbildung viel häufiger die Städte
während des Krieges verlassen haben, als die (späteren)
Volksschulabsolventen. Vermutlich kommen hier die jeweils unter-
schiedlichen finanziellen Ressourcen der Eltern und je nach Schul-
typ differierende, staatliche Umsiedelungsprogramme (Kinderlandver-
schickung) zum Tragen.
Nur bei den Realschul- und Gymnasialabsolventen nimmt der Anteil
der Städter ab dem 25. Lebensjahr wieder ab. Ob es sich hierbei um
Rückwanderungen in die ländlichen Herkunftsorte oder eher um Abwan-
derungen in Randgemeinden von Großstädten handelt, ist nicht
entscheidbar.
Für alle drei Kohorten gilt, daß ab dem 16./17. Lebensjahr der An-
teil von Bewohnern ländlicher Regionen abnimmt. Ein Teil der Wande-
rungen in die Städte wird ausbildungsbedingt sein, offen ist aber,
in welchem Ausmaß solche Wanderungen vorgenommen werden und ob sich
hierbei - möglicherweise als Folge einer Dezentralisierung von Bil-
dungseinrichtungen - bedeutsame kohortenspezifische Unterschiede
zeigen. Tabelle 1 gibt die Verteilung bildungsmotivierter Wande-
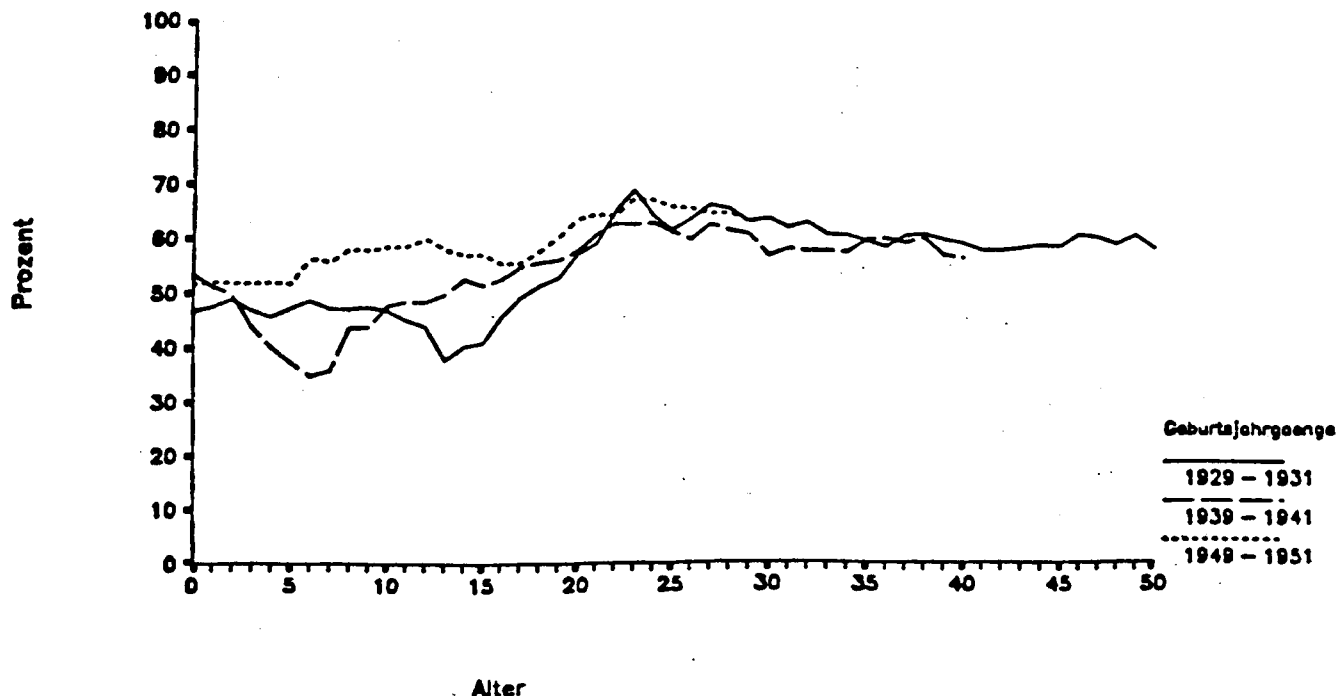rungen nach Geburtsjahrgängen wieder.

Abbildung 1

Personen mit niedrigem Schulabschluss
(Volksschule oder weniger) in Gross— oder Mittelstaedten

nach Kohortenzugehoerigkeit

Alter
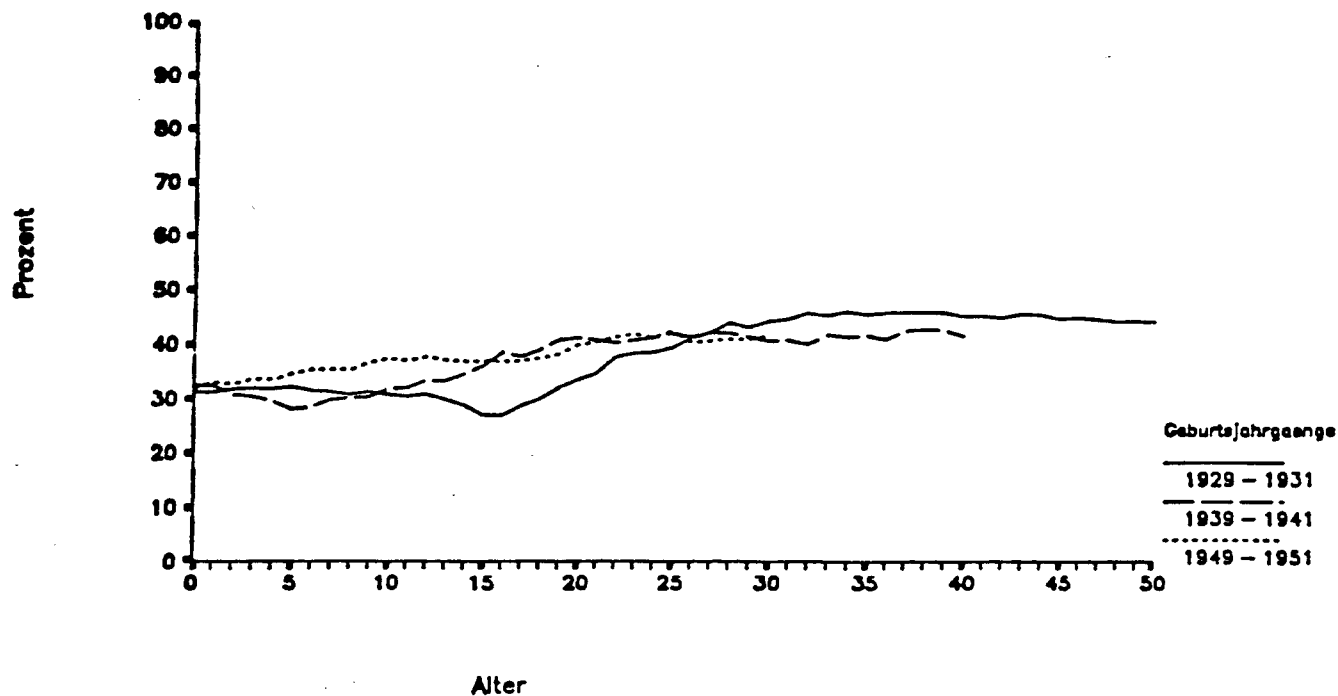
Quelle: Lebensverlaufsstudie des Sfb—3

Abbildung 2

Personen mit hohem Schulabschluss
(Realschule oder Abitur) in Gross— oder Mittelstaedten

nach Kohortenzugehoerigkeit

Alter

Quelle: Lebensverlaufsstudie des Sfb—3

Tabelle 1

Migrationen mit Ausbildungsmotiv (in %), nach Alter,
           Geburtsjahr und Geschlecht

Alter	Geburtsjahr			Männer	Frauen
	1929-1931	1939-1941	1949-1951		
0-14	6.6	5.2	3.3	5.3	4.7
15-19	7.6	9.9	10.5	7.5	11.2
20-24	6.4	8.4	10.0	11.1	5.4
25 u.m.	2.1	2.9	4.0	4.1	1.8
Personen (= 100 %)	707	728	733	1087	1081

*) Ein Ausbildungsmotiv lag vor, wenn als Wanderungsmotiv
   "Schule", "Ausbildung", "Studium", "Lehre", "Internat",
   "Weiterbildung" oder "Meisterprüfung" angegeben wurde.

Sieht man von den Wanderungen ab, die vor dem 15. Lebensjahr vorge-
nommen wurden, dann zeigt sich in der Abfolge der Kohorten eine
schwache Zunahme bildungsmotivierter Migrationen. Am häufigsten
werden sie am Ende des zweiten Lebensjahrzehnts durchgeführt,
zunehmend jedoch auch in nachfolgenden Altersgruppen. Es ist somit
klar, daß das Gros dieser Wanderungen eine Berufsausbildung und
nicht eine weiterführende, allgemeine Schulausbildung ermöglichen
soll. Frauen nehmen aufgrund ihrer kürzeren Bildungsdauer ausbil-
dungsorientierte Wanderungen früher als Männer vor.
Die Bildungsbeteiligung hat in der Kohortenabfolge vor allem bei
den Frauen deutlich zugenommen (47,3 % der Frauen in der ältesten
Kohorte begannen eine Berufsausbildung, 65,8 % waren es bereits
beim Geburtsjahrgang 1939-1941 und 83,1 % bei der jüngsten
Kohorte), dem steht jedoch kein vergleichbarer Trend bei den aus-
bildungsmotivierten Wanderungen gegenüber. Es ist festzustellen,
daß trotz einer möglichen Verringerung regionaler Ungleichheiten im
Bildungsbereich immer noch eine bemerkenswerte Anzahl von Wande-

rungen zu Ausbildungszwecken vorgenommen wird. So war vermutlich
insbesondere die Kohorte 1949-51 von dem starken Rückgang der
betrieblichen Ausbildungsplätze zwischen 1970 und 1975 betroffen
(DERENBACH 1984: 21 f.). Da es sich dabei wahrscheinlich um einen
regional unterschiedlich verlaufenden Prozeß gehandelt hat, könnte
dieses ein Grund für die leichte Zunahme ausbildungsmotivierter
Wanderungen zwischen der mittleren und jüngsten Kohorte sein. Es
gab und es gibt Bevölkerungsgruppen, die durch eine schlechte in-
frastrukturelle Ausstattung mit Bildungseinrichtungen benachteiligt
waren bzw. sind. Über Veränderungen dieser regionalen Disparitäten
lassen sich aus den obigen Daten keine Schlüsse ziehen, da über
wichtige Faktoren, die zu einer bildungsbedingten Wanderung führen
können - wie Bildungsaspirationen, Wohnungsmarktfaktoren, finan-
zielle Lage - keine Daten vorliegen.
Im theoretischen Teil wurde die These aufgestellt, daß Personen,
die aus ländlichen Regionen stammen, aufgrund regionaler Dispari-
täten im Bildungswesen besonders häufig ausbildungsbedingte Wande-
rungen vornehmen. Diese Auffassung kann mit dem vorhandenen Daten-
material nicht generell gestützt werden. Nach Tabelle 2, die an-
gibt, wie viele Wanderungen mit einem Ausbildungsmotiv auf 100 Per-
sonen entfallen, die jeweils Orten unterschiedlicher Größe entstam-
men, zeigt sich keine lineare Beziehung. Am häufigsten nehmen Per-
sonen, die aus Kleinstädten stammen, Wanderungen aus Ausbildungs-
gründen vor (33 Wanderungen per 100 Personen), diese Rate nimmt
dann mit zunehmender Größe des Herkunftsortes ab - insofern wird
die obige These bestätigt. Doch diejenigen, deren erster Wohnort im
Lebenslauf ein Dorf war, weisen weniger ausbildungsorientierte
Wanderungen auf als die Klein- und Mittelstädter. Wenn man unter-
stellt, daß die infrastrukturelle Ausstattung mit Bildungsein-
richtungen in dörflichen Gemeinden am schlechtesten ist, müßten die
bildungsmotivierten Wanderungen bei ihren Bewohnern am ehesten
auftreten. Da dieses nicht der Fall ist, kann man vermuten, daß
Jugendliche aus dörflichen Gemeinden in hohem Ausmaß örtlich
gebunden sind, geringere Bildungsaspirationen aufweisen oder
weniger Informationen über Bildungseinrichtungen in anderen Orten
besitzen. Nur genauere Analysen könnten zeigen, welche dieser
Erklärungen zutreffend ist.

Tabelle 2

Anzahl der Migrationen mit Ausbildungsmotiv (in %) bis zum 30. Lebensjahr, nach Geburtsjahrgang und regionaler Herkunft

Alter	Größe des ersten Wohnortes			
	Dorf	Kleinstadt	Mittelstadt	Großstadt
1929-1931	13.1	34.1	30.3	27.0
	(335)	(132)	(89)	(152)
1939-1941	23.6	25.4	38.3	25.3
	(334)	(126)	(81)	(186)
1949-1951	22.3	38.7	27.0	24.6
	(301)	(150)	(111)	(171)
Gesamt	19.6	33.1	31.3	25.5
	(970)	(408)	(281)	(509)

*) Die Werte in Klammern geben die Anzahl der Personen pro
   Subgruppe (= Prozentuierungsbasis) an.

Deutliche Kohortentrends werden auch bei einer Differenzierung
nach der regionalen Herkunft nicht sichtbar. Als Einzelergebnis ist
interessant, daß Personen des Geburtsjahrgangs 1929-1931, die in
Dörfern aufgewachsen sind, nur wenig in ihre Ausbildung qua Migration investiert haben - eine Folge des Krieges: In den Großstädten
war an eine berufliche Ausbildung nicht zu denken.

4.2 Alters- und kohortenspezifische räumliche Mobilität bei unter-
   schiedlichen Bildungsgruppen

Tabelle 3 zeigt altersspezifische Migrationsraten nach der Kohor-
tenzugehörigkeit bei Personen, die höchstens einen Volksschulab-
schluß haben, gegenüber denen, die eine Realschule oder ein Gymna-
sium absolviert haben.
In nahezu allen Altersgruppen sind Personen mit Realschulabschluß
oder Abitur räumlich mobiler als diejenigen mit einem niedrigeren
Bildungsniveau. Zum Teil sind die Raten doppelt so hoch: Bei der
mittleren Kohorte entfallen auf 100 Volksschulabsolventen 55 Wande-
rungen bei den 25- bis 29jährigen, bei der höheren Bildungsgruppe
sind es 102. Die einzige Ausnahme bilden die Angehörigen der
mittleren und jüngsten Kohorte, die zwischen 15 und 20 Jahre alt
sind. Hier kommt der Umstand zum Tragen, daß sich die Volksschulab-
solventen im Gegensatz zu Angehörigen der zweiten Bildungsgruppe zu
einem größeren Teil nicht mehr im allgemeinbildenden Schulsystem
befinden, was sich generell als mobilitätsfördernd auswirken wird.
Bemerkenswert ist weiterhin, daß auch bei einer Differenzierung
nach Bildungsgruppen die bekannte Form der Altersselektivität von
Wanderungen erhalten bleibt: In beiden Bildungsgruppen sind die 20-
bis 24jährigen am mobilsten und wechseln mit zunehmendem Alter dann
seltener ihre Wohnung. Eine Abnahme der Bildungsselektivität mit
dem Alter läßt sich jedoch nicht nachweisen. Das Verhältnis der
Migrationsraten bleibt in der Abfolge der Altersgruppen eher
stabil. Demzufolge kommt der Bildung ein Effekt auf räumliche
Mobilität zu, der nicht auf frühe Phasen des Lebenslaufs beschränkt
bleibt.
Dem Interkohortenvergleich kann man entnehmen, daß Realschulabsol-
venten und Abiturienten der mittleren Kohorte in allen vergleich-
baren Altersgruppen geographisch besonders mobil waren.
Auch die Differenzen der Migrationsraten bei den Bildungsgruppen
sind hier relativ hoch, die Bildungsselektivität räumlicher Mobili-
tät ist somit bei den um 1940 Geborenen am höchsten, bei der
jüngsten Kohorte dagegen am niedrigsten. Für diese Unterschiede
sind sicherlich Bedingungen der Kriegs- und unmittelbaren

Nachkriegsjahre verantwortlich zu machen, deren nähere Untersuchung
hier jedoch nicht erfolgen kann. Ferner ist davon auszugehen, daß
nach höherqualifizierten Arbeitskräften zu Beginn der 60er Jahre
eine besonders große Nachfrage - vor allem in den Städten - be-
stand, so daß durch berufsorientierte Migrationen relativ hohe Ein-
kommenssteigerungen erzielt werden konnten. Außerdem waren die 60er
Jahre davon geprägt, daß sich die Zahl der Betriebsstandorte
erhöhte (DERENBACH 1984: 88). Kann der Arbeitskräftebedarf nach
weniger qualifizierten Arbeitskräften häufig noch "vor Ort" gedeckt
werden kann, so ist dieses bei Höherqualifizierten seltener der
Fall.

Tabelle 3

Altersspezifische Migrationsraten (in %), nach Geburtsjahrgang
und Schulabschluß

| | Geburtsjahrgang | | | | | |
| Altersklasse | 1929-1931 | | 1939-1941 | | 1949-1951 | |
	VS	MS/ABI	VS	MS/ABI	VS	MS/ABI
15-19	68.6	86.9	56.1	51.3	49.8	47.4
20-24	73.0	100.8	82.4	126.0	89.3	112.1
25-29	59.1	96.1	55.0	102.6	56.5	87.1
30-34	35.3	48.5	32.7	59.1	--	--
35-39	22.5	35.4	20.1	35.7	--	--
40-44	17.3	23.1	--	--	--	--
45-49	10.7	14.6	--	--	--	--
Anzahl Per-sonen (= 100%)	577	130	574	154	494	239

*) Unterschieden werden Personen mit höchstens einem Volks-
schulabschluß (VS), Mittlerer Reife (MS) oder Fachhochschul-
reife bzw. Abitur (ABI). Angegeben wird die Anzahl der Wande-
rungen pro 100 Personen der jeweiligen Subgruppe.

4.3 Der Effekt des Bildungsniveaus auf Nah- und Fernwanderungen

Daß Migrationsraten zwischen den Bildungsgruppen variieren, hat die
bisherige Untersuchung deutlich gemacht. Wir wollen uns nun mit der
Frage befassen, in welchem Ausmaß räumliche Mobilität vom Bildungs-
niveau abhängt und ob diese Bildungsselektivität möglicherweise
ein Phänomen ist, daß sich nur bei bestimmten Arten von Wanderungen
zeigt. Darüber hinaus stellt sich das Problem, warum das Bildungs-
niveau einen Effekt auf räumliche Mobilität hat. Zu diesem Problem
wurden bereits einige Thesen formuliert, die mit empirischen Ergeb-
nissen konfrontiert werden sollen.

Aus Tabelle 4, die sich nur auf Erwerbstätige bezieht, geht hervor,
daß das Ausmaß des Einflusses von Bildung auf räumliche Mobilität
davon abhängt, ob die Effekte anderer Faktoren, insbesondere des
Alters, der beruflichen Stellung und der Ortsbindung, berücksich-
tigt werden. Was die Nahwanderungen anbelangt, so wird verständ-
lich, warum die bisherige Forschung zu gegensätzlichen Ergebnissen
kam: Der Zusammenhang zwischen Bildung und Nahwanderungen wird erst
dann sichtbar, wenn er altersspezifisch untersucht wird. Auch
Nahwanderungen werden eher von Personen mit einem höheren Bildungs-
niveau vorgenommen.

Das Bildungsniveau hat indessen auf Fernwanderungen einen deutlich
stärkeren Effekt als auf Nahwanderungen, an Wanderungsströmen über
größere Distanzen sind Höherqualifizierte überproportional betei-
ligt. Betrachten wir die letzte Zeile in Tabelle 4, so ist das
Risiko eines Wohnungswechsels über eine kürzere Distanz für Real-
schulabsolventen um 16 % höher als bei Personen mit einem Volks-
schulabschluß ((exp (0.15)-1) * 100 %) = 16 %), Abiturienten weisen
eine Rate für Nahwanderungen auf, die jene der Volksschulabsolven-
ten um 51 % übersteigt, für Fernwanderungen lauten die analogen
Werte 46 % und 166 %.

Offensichtlich wird ein Teil des Zusammenhangs zwischen Bildung und
Fernwanderungen durch die berufliche Stellung der Individuen
erklärt. Aber es bleibt ein starker Bildungseffekt bestehen.
Dasselbe gilt für die Ortsbindung: Sie reduziert den Bildungsef-
fekt, ohne ihn aufzuheben. Insofern treffen Hypothesen zu, wonach
Höherqualifizierte deshalb eher Fernwanderungen vornehmen, weil sie
Positionen auf dem Arbeitsmarkt einnehmen, die dieses erfordern
oder sie aufgrund vorangegangener Wanderungen weniger enge lokale
Beziehungen aufweisen. Doch diese Faktoren reduzieren den Bildungs-
effekt nur in einem geringen Ausmaß.

Tabelle 4

Effekte des Bildungsniveaus auf Nah- und Fernwanderungen *)

	Distanz			
Kontroll-	Nah		Fern	
variablen	MS	ABI	MS	ABI
KEINE	0.07	0.12	0.36***	0.77***
KOHO	0.05	0.12	0.37***	0.77***
KOHO,SEX	0.04	0.13	0.35***	0.79***
KOHO,SEX,ALTER	0.13**	0.37***	0.57***	1.25***
KOHO,SEX,ALTER, BST	0.17**	0.46***	0.45***	1.15***
KOHO,SEX,ALTER, BST,OBDG	0.15**	0.41***	0.38***	0.98***

Anzahl			
-Wohnperioden	1698		619
-Zensierungen	2012		3091

*) Datenbasis sind allen Wohnperioden, bei deren Beginn die Personen erwerbstätig waren; Vergleichsgruppe sind die Volksschulabsolventen.

Codierungen:

MS: Mittlere Reife; ABI: Fachhochschulreife, Abitur;

KOHO: Kohorte - 1929-31 (0/1); 1939-41 (0/1); 1949-51 (0/1);

SEX: Geschlecht - 0: Männer; 1: Frauen;

ALTER: Alter bei Wohnbeginn - fünf Altersklassen, jeweils mit 0/1 codiert;

BST: berufliche Stellung - jeweils zwei mit 0/1 codierte Kategorien von Arbeitern, Angestellten und Beamten sowie Selbständige und andere;

OBDG: Ortsbindung - 0: Person lebt nicht im Geburtsort; 1: Person lebt im Geburtsort;

Signifikanzniveau: * - 10%; ** - 5%; *** - 1%;

4.4 Bildungsselektive Migrationen in Abhängigkeit von dem Verdich-
    tungsgrad der Herkunfts- und Zielregion

In einem letzten Schritt möchte ich untersuchen, ob die schulische
Ausbildung auf die Richtungen der Wanderungen Einfluß nimmt. Tabel-
le 5 zeigt auf der Basis der Klassifikation, die in Kap. 3 er-
läutert wurde, daß die Größe von Herkunfts- oder Zielregion ein
bedeutsames Kriterium ist, wenn es darum geht zu präzisieren, wie
sich die Bevölkerung nach ihrem Bildungsniveau regional differen-
ziert.
Ohne nun im einzelnen auf die Variation von Bildungseffekten einzu-
gehen, die durch die Einführung weiterer Faktoren entsteht, ist es
eindeutig, daß Fernwanderungen aus ländlichen in städtische Regio-
nen die höchste Bildungsselektivität aufweisen. Dabei sollte aller-
dings nicht übersehen werden, daß es einen, wenn auch schwächeren,
Gegenstrom gibt. Doch nur die Abiturienten, nicht die Absolventen
von Realschulen, verlegen ihren Wohnsitz überproportional häufig
von der Stadt auf das Land.
Die Attraktivität städtischer Regionen für die Höherqualifizierten
kommt ferner klar darin zum Ausdruck, daß sich die Bildungsgruppen
bei Fernwanderungen zwischen Städten stark voneinander unterschei-
den. Wir können daraus schließen, daß nicht so sehr die Größe des
Herkunftsgebietes als vielmehr die Größe der Zielregion der
entscheidende Grund für den Einfluß des Bildungsniveaus auf Migra-
tionen sind. So sind an Fernwanderungen zwischen ländlichen Regio-
nen Volksschulabsolventen ähnlich häufig beteiligt wie Realschulab-
solventen oder Abiturienten, es kommt hier zu keiner
Bildungsselektivität.
Gehen wir nun zu den Nahwanderungen über, so lassen sich - wenn
überhaupt - nur schwache Bildungseffekte beobachten. Verlaufen sie
von einer ländlichen in eine städtische Region, so unterscheiden
sich Personen nach ihrem Schulabschluß nicht. Innerhalb ländlicher
Räume zeigt sich ein schwacher, nicht-linearer Zusammenhang: Nur
Realschulabsolventen unterscheiden sich von der niedrigsten Bil-
dungsgruppe, nicht jedoch die Abiturienten. Bei Wohnungswechseln
innerhalb von Städten ist die Bildungsselektivität gering, führen
sie in städtisches Umland, spielt die Schulbildung keine Rolle.

Tabelle 5

Effekte der Bildung auf Migrationen nach ihrer Distanz und Richtung *)

Bildungs-niveau	Distanz							
	Nah				Fern			
	L/S	L/L	S/L	S/S	L/S	L/L	S/L	S/S
-- ohne Kontrollvariablen --								
MS	0.02	0.17	0.07	-0.04	0.56**	-0.40	0.03	0.70***
ABI	0.31	-0.11	0.46	0.08	1.15***	0.24	0.79***	0.82***
-- unter Kontrolle von Kohortenzugehörigkeit, Geschlecht, Alter, beruflicher Stellung und Ortsbindung --								
MS	-0.01	0.28**	-0.02	0.02	0.59**	-0.28	0.01	0.67***
ABI	0.29	0.14	0.60	0.43***	1.36***	0.51	0.95***	1.15***
	(114)	(682)	(118)	(781)	(137)	(145)	(132)	(193)

*) L: Weiler, Dorf, Kleinstadt;   S: Mittel- oder Großstadt; Herkunftsregion/ Zielregion.

Die Werte in Klammern geben die Anzahl der Ereignisse pro Modell an.

Weitere Erläuterungen siehe Tabelle 4

5. Schlußfolgerungen

Das Bildungsniveau ist eine Determinante räumlicher Mobilität, der eine bedeutsame Erklärungskraft zukommt, auch dann, wenn das Alter, die berufliche Stellung oder lokale Bindungen in der Analyse berücksichtigt werden.

Die Konzentration der Höherqualifizierten in den größeren Städten
ist Resultat

- einer bildungsselektiven Land-Stadt-Wanderung, deren Folgen für
  die Sozialstruktur der Zielregion nicht durch einen analogen
  Wanderungsstrom in die Gegenrichtung ausgeglichen werden,
- des hohen Anteils von Realschul- und Gymnasialabsolventen, die
  in größeren Städten aufgewachsen sind.

Die Bildungsbeteiligung hat sich kräftig erhöht, ausbildungsmoti-
vierte Wanderungen haben dagegen nur schwach zugenommen. Immerhin
können sie quantitativ nicht vernachlässigt werden, und in die in-
dividuellen Kosten von Ausbildungen sind nicht nur Einkommensver-
luste einzubeziehen, sondern auch monetäre und nicht-monetäre
Wanderungskosten. Diese Kosten werden vor allem von Personen aufge-
bracht, die in Orten mittlerer Größe aufgewachsen sind. Es bedarf
weiterer Untersuchungen zu der Frage, warum Dorfbewohner seltener
Wanderungen zu Ausbildungszwecken vornehmen als Personen, die aus
Klein- und Mittelstädten stammen. Durch die geringen "Bildungsmi-
grationen" der Großstädter kommt es zu einem U-förmigen Zusammen-
hang, der durch die Wirkungen regionaler Disparitäten im Bildungs-
bereich allein kaum erklärt werden kann. Der Arbeitsmarkt der
Höherqualifizierten ist nicht nur regional weit ausgedehnt, sondern
er ist zudem in den Verdichtungsräumen zu verorten. Dieser Tatbe-
stand ist es, der bildungsselektive Migrationen bedingt, der aber
auch ein Grund dafür sein dürfte, daß qualifizierte Ausbildungen in
erster Linie in den größeren Städten absolviert werden können.

Die Bildungsselektivität der Fernwanderungen ist deutlich belegbar,
sie ist jedoch abhängig von der Größe der Zielregion. Es sind die
Wanderungsströme in die Städte, mit ländlichen oder städtischen
Herkunftsregionen, an denen die Höherqualifizierten überproportio-
nal beteiligt sind. Damit haben bildungsselektive Wanderungen in
erster Linie Folgen für die Sozialstruktur ländlicher Regionen.
Dieses wiederum wird einen bedeutsamen Effekt auf die betriebliche
Standortwahl haben: Da in ländlichen Regionen das Angebot an quali-
fizierten Erwerbstätigen gering sein dürfte, siedeln sich viele
Unternehmen dort nicht an. Auch die Steigerung des Bildungsniveaus
von Bewohnern ländlicher Regionen wird zunächst eher deren Abwande-
rung verstärken, als der lokalen Wirtschaft zugute kommen.
Offensichtlich sind die Wohnungswechsel der Höherqualifizierten ein
wichtiger Teilprozeß bei der räumlichen Allokation von

Produktionsfaktoren. Daher existieren gesellschaftliche Regelungen und Normen, deren Funktion es ist, eine im ökonomischen Sinn genügend hohe räumliche Mobilität von "Humankapital" zu gewährleisten. Dazu gehören Kommunikationssysteme, die eine überregionale Rekrutierung von Arbeitskräften ermöglichen (z.B. Stellenanzeigen in Zeitungen), staatliche Umzugshilfen (Berlinförderung, Steuererleichterungen) und eine Ideologie des modernen Menschen, der immer auch räumlich mobil sein soll.

6. Literatur

ALBRECHT, G., 1972: Soziologie der geographischen Mobilität. Stuttgart.

BECKER, G.S., 1964: Human capital. New York, London.

BECKER, H.S., 1960: Notes on the concept of commitment. American Journal of Sociology 66, 32-40.

BIRG, H., 1985: Interregionale demo-ökonomische Modelle für die Bundesrepublik Deutschland: Eine Zwischenbilanz. Institut für Bevölkerungsforschung und Sozialpolitik, Universität Bielefeld, IBS-Materialien Nr.18

BLOSSFELD, H.-P., 1985a: Zur Repräsentativität der Sfb3-Lebensverlaufsstudie - ein Vergleich mit Daten aus der amtlichen Statistik. Sfb3-Arbeitspapier Nr. 163. Frankfurt a.M., Mannheim.

---, 1985b: Bildungsexpansion und Berufschancen. Frankfurt a.M., New York.

BLOSSFELD, H.P., HAMERLE, A. & MAYER, K.U., 1986: Ereignisanalyse. Frankfurt a.M., New York.

BOGUE, D.J. & HAGOOD, M.J., 1953: Differential migration in the Corn and Cotton Belts. Subregional migration in the United States, 1939-40, Vol. II, Oxford, Ohio.

BRÜCKNER, E. u.a., 1984: Methodenbericht "Lebensverläufe". ZUMA-Technischer Bericht Nr. T 84/08. Mannheim.

COURGEAU, D., 1984: Relations entre cycle de vie et migrations. Population 39, 483-514.

---, 1985: Interaction between spatial mobility, family and career life-cycle: a French survey. European Sociological Review 1, 2, 139-162.

COX, D.R., 1972: Regression models and life-tables (with discussion). Journal of the Royal Statistical Society B, 34, 187-220.

---, 1975: Partial likelihood. Biometrika 62, 269-276.

DERENBACH, R., 1984: Berufliche Eingliederung der nachwachsenden Generation. Forschungen zur Raumentwicklung Bd. 13. Bonn.

FORD, D.A., 1973: Commitment: a mathematical model. Quality and Quantity 7, 1-40.

FRIEDLANDER, D. & ROSHIER, R.J., 1966: A study of internal migration in England and Wales. (Part II). Population Studies 20, 1, 45-59.

GEIPEL, R., 1965: Sozialräumliche Strukturen des Bildungswesens. Frankfurt a.M. u.a.

GRUBEL, H.G. & SCOTT, A., 1967: Determinants of migration: the highly skilled. International Migration 5, 127-139.

HEBERLE, R., 1936: Die Bedeutungen der Wanderungen im sozialen Leben der Völker. In: ALBRECHT u.a. (Hg.): Reine und angewandte Soziologie. Eine Festgabe für Ferdinand Tönnies. Leipzig.

HOFBAUER, H. & NAGEL, E., 1973: Regionale Mobilität bei männlichen Erwerbspersonen in der Bundesrepublik Deutschland. Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 6, 255-272.

HORSTMANN, K., 1976: Zur Soziologie der Wanderungen. In: KÖNIG, R. (Hg.): Handbuch der empirischen Sozialforschung, Bd.5, Soziale Schichtung und Mobilität. Stuttgart.

KIERNAN, K., 1977: Characteristics of young people who move interregionally: a longitudinal study. In: HOBCRAFT, J. & REES, P. (Hg.): Regional demographic development. London.

LANSING, J.B. & MUELLER, E., 1967: The geographic mobility of labor. Ann Arbor, Michigan.

LONG, L.H., 1973: Migration differentials by education and occupation: trends and variations. Demography 10, 243-258.

MAMMEY, K., 1979: Räumliche Aspekte der sozialen Mobilität in der Bundesrepublik Deutschland. In: TEGTMEYER, H. (Hg.): Soziale Strukturen und individuelle Mobilität. Wiesbaden.

MAREL, K., 1980: Inter- und intraregionale Mobilität. Wiesbaden.

MORRISON, P.A., 1972: Population movements and the shape of urban growth: implications for public policy. Santa Monica.

MUSGROVE, F., 1963: The migratory elite. London.

NIEDZWETZKI, K., 1979: Die Wanderungen: Räumliche Mobilität und Regionalpolitik. Raumordnungsbericht Bd. 12, Schwäbisch Gmünd.

PAPASTEFANOU, G.: 1980: Zur Güte von retrospektiven Daten. Eine Anwendung gedächtnispsychologischer Theorie und Ergebnisse einer Nachbefragung. Sfb-3 Arbeitspapier Nr. 29, Frankfurt a.M., Mannheim.

PEISERT, H., 1967: Soziale Lage und Bildungschancen in Deutschland. München.

QUIGLEY, J.M. & WEINBERG, D.H., 1977: Intra-urban residential mobility: a review and synthesis. International Regional Science Review 2, 41-66.

ROB 1978 (Raumordnungsbericht der Bundesregierung), hg. vom Bundesminister für Raumordnung, Bauwesen und Städtebau. Bonn.

ROB 1982 (Raumordnungsbericht der Bundesregierung), hg. vom Bundesminister für Raumordnung, Bauwesen und Städtebau. Bonn.

ROB 1986 (Raumordnungsbericht der Bundesregierung), hg. vom Bundesminister für Raumordnung, Bauwesen und Städtebau. Bonn.

SANDEFUR, G.D. & SCOTT, W.J., 1981: A dynamic analysis of migration: an assessment of the effects of age, family and career variables. Demography 18, 3, 355-368.

SCHREIBER, K.H., 1975: Wanderungsursachen und idealtypische Verhaltensmuster mobiler Bevölkerungsgruppen. Frankfurt a.M.

SHAW, R.P., 1975: Migration theory and fact. Philadelphia, Pennsylvania.

SHYROCK, H.S. & NAM, C.B., 1965: Educational selectivity of interregional migration. Social Forces, 3, 299-310.

SUVAL, E.M. & HAMILTON, C.H., 1964: Some new evidence on educational selectivity in migration to and from the South. Social Forces 43, 536-547.

TÖLKE, A, 1982: Möglichkeiten und Grenzen einer Edition bei retrospektiven Verlaufsdaten: Dargestellt an der Lebensver-laufsbefragung des Sfb3. Sfb3-Arbeitspapier Nr. 90, Frankfurt a.M., Mannheim.

TUMA, N.B. & HANNAN, M.T., 1984: Social dynamics. Orlando u.a.

WIEDENBECK, M., 1982: Zum Problem repräsentativer Querschnitte von kleinen Teilgruppen der Bevölkerung am Beispiel des Projekts "Lebensverläufe und Wohlfahrstentwicklung". ZUMA-Nachrichten 10, 21-34.

Post-war fertility patterns in the Federal Republic of Germany

Nancy B. Tuma and Johannes Huinink

Since the late nineteenth century, the overall level of
fertility in Germany has tended to decline, as it has in all
industrialized nations. Especially since the mid-1960s,
however, it has declined even more strikingly until the net
reproduction rate in the Federal Republic of Germany was
only 0.605 in 1984: 0.603 for the native German population
and 0.659 for the foreign population, which is mainly of
Turkish and Southern European origin.[1] One can easily
understand why such figures have led to the popularization
of the question of "whether Germans are becoming extinct"
and to renewed discussion of various pro-natalist policies.

Our goal is neither to recommend population policies
nor to forecast the consequences of recent fertility trends.
Instead, we describe fertility patterns of native Germans in
the post-war era in some detail and discuss the relationship
of these patterns to factors that may have promoted them.

Previous investigations of fertility trends in the
Federal Republic of Germany have mainly fallen into one of
several types. The predominant type has been some form of

demographic analysis based on aggregate data. Recently
there have been some cohort analyses, but again using
aggregate data (Dinkel, 1983; Birg et al., 1984). The main
shortcoming of these studies, largely due to the inherent
limitations of official statistics, is their inability to
examine heterogeneity in fertility patterns within the
German population (e.g., due to birth order or background
variables) or to study change over time from an individual
perspective.

Another common way of studying German fertility has
been to analyze individual-level, cross-sectional data on
the desired number of children. Relative to the analyses
mentioned above, such studies have the advantage of
revealing population heterogeneity in a key indicator of
fertility and of showing the relationship between this
indicator and various individual attributes (for an
overview, see Kiefl and Schmid, 1985). But they typically
do not consider the relationship between the desired number
of children and actual fertility, let alone the dynamics of
fertility from an individual perspective.

There have been several recent attempts to remedy some
of these deficiencies, which are by no means limited to
analyses of German fertility. One line of work has tried to
relate processes of change on the societal level to changes
in family structure and fertility (Huinink, 1987). Another
approach has tried to analyze the dynamics of individual
behavior, which is treated as the outcome of a set of
interdependent choices in different life domains, such as
schooling, employment, and marriage (for example, see
Michael and Tuma, 1985; Strohmeier, 1985). These choices
are viewed as affected by a person's background and current
life situation. Though highly individualized, this approach
need not disregard historical circumstances and experiences,
especially when different birth cohorts can be compared. It
is the approach we adopt in this research.

Our research on fertility in West Germany is divided
into two main stages. The first is an investigation of the

overall pattern of fertility ignoring all covariates except
a person's sex and birth cohort. In this first stage we
examine survival probabilities, age-specific transition
rates, and, for higher order births, duration-specific
transition rates. The goal here is primarily to describe
basic patterns of fertility of native German men and women
in different birth cohorts, but in more detail than
aggregate data permit. In the second stage we estimate
models of birth rates at different parities as a function of
selected covariates that describe a person's social
situation and that are hypothesized to affect these rates.
This paper reports findings of the first stage of research;
results of the second stage are given in a forthcoming
companion paper.

THE DATA

Analyses of individual dynamics of fertility require a
fairly unusual form of data: longitudinal data on fertility
of a sample of individuals with known characteristics. The
data we analyze were gathered in the 1981-1982 German Life
History Survey (GLHS) sponsored by Sonderforschungsbereich 3
and the Max-Planck-Institut fuer Bildungsforschung under the
leadership of Karl Ulrich Mayer.[2] The respondents were
selected through a sampling plan that was designed to give a
representative picture of native Germans living in the
Federal Republic of Germany in 1981-1982 who had been born
in one of three periods: 1929-1931, 1939-1941, and
1949-1951.[3] We refer to these as the 1930, 1940, and 1950
cohorts for short. The total sample size is 2171, roughly
half men and half women.

Each respondent was personally interviewed using a
questionnaire that focused especially on the individual's
previous life history. It collected detailed information on
both parents and siblings of respondents, and on their own
schooling, vocational training, employment, residential
history, marital history, and fertility history. Limited

information on the current spouse was also collected for respondents who were married at the interview.

These data are unusual in several respects. First, they give the dates of events (usually to the nearest month and year), and not just a picture of the respondent at specific ages or in specific years. That is, the data provide event histories for several major life domains. Second, they contain information on a much wider range of life domains than do most other data sets, whether on West Germans or other populations. Third, by focusing on three cohorts, these data let one study change over time in the behavioral patterns of West Germans in the post-war era.

These data are not without limitations, of course. Three are worth stressing. First, the sample is much smaller than is ideal when investigating changes across cohorts and differences between men and women. Second, the particular cohorts included in the sample mean that there are essentially no data for studying cohort differences in fertility in several interesting historical periods, in particular, the economic boom in the late 1960s. (The youngest cohort is too young to have many children during this period, and the oldest cohort had largely completed its child-bearing before it began.) Third, roughly eight percent of the West German population comes from countries where fertility tends to be higher than in West Germany. Because of the sample design, the GLHS data cannot be used to compare native Germans and the nonnative subpopulations. Moreover, because the GLHS sample omits nonnatives, care must be taken in extrapolating from these data to the Federal Republic of Germany as a whole.

RESULTS

Description of the Sample. Before discussing fertility of the men and women responding to the 1981-1982 German Life History Survey, we describe a few of their characteristics

that may be related to their fertility. Means of selected variables are given in Table 1.

We report the person's education and participation in vocational training because these variables are excellent predictors of future life chances and may also be associated with family-related values. (For example, less educated individuals are likely to be more family-oriented and to have more children on average.) There is a marked shift across the cohorts to higher levels of education and, for women, more training. Still, women tend to be less educated than men, even in the youngest cohort, and they are much less likely than men to have had vocational training.

Over time the structure of respondents' parental homes has changed substantially. In particular, the average family size of respondents' parents declines across the cohorts. If a person's fertility is partly imitative of his or her parents', then we expect those with more siblings to tend to have more children. Hence we expect some decline in respondents' fertility across the three cohorts. The intercohort trend in respondents' average number of children does parallel that of the parents', but is markedly lower.

We also report a few other aspects of the background of the respondents since the typical characteristics of the West German population may not be familiar. The fraction raised as Catholics is roughly one-half for men and women in all three cohorts. The increasing urbanization of the German population is revealed by the rising proportion living in large places at age 14 and the falling proportion living in small places. The fraction living with only one parent at age 14 is under one-tenth, except for the 1940 cohort, for whom it rises to roughly one-fifth, undoubtedly due to World War II.

Table 1 also reports the mean age difference (in years) between the respondent and his/her spouse for those who were married at the interview. It is about three years for all three cohorts.

Number of Children Born to Respondents. Table 2
reports the number of children born to the respondents
before the interview by sex and cohort in two ways. Panel A
reports the number who have had exactly each number of
children by the interview, as well as the mean (see also
Table 1); Panel B gives the number who have had at least a
given number of children (i.e., the number who have ever
been at risk of having the next higher number of children).

The entries in Panel A correspond to completed family
size for the 1930 cohort, whose members were over 50 years
old at the interview. Since relatively few children are
born after age 40 to women (and also, it appears, to men),
the entries in Panel A probably underestimate completed
family size for the 1940 cohort only slightly (i.e., by
under 0.02 children). But those in the youngest cohort
ranged from 30 to 33 years old at the interview and can be
expected to have roughly 0.25 additional children on average
before completing their fertility. Thus, the entries in
Panel A are not a good indicator of completed family size
for the 1950 cohort. Still, these figures suggest that
completed family size has declined across these cohorts.

It is interesting that the maximum number of children
born to any respondent is eight. In contrast, family size
of respondents' parents is five or more for 15 percent of
the sample and eight or more for three percent. Thus, not
only has average family size has shrunk considerably across
the generations, but large families have become much rarer.

Panel B of Table 2 shows that, however much one might
like to study birth rates at higher parities, it is not
possible with these data. The number of births of order
four or more is simply too few to support such an analysis.
Therefore, we concentrate on the first birth, the transition
from the first birth to the second, and the transition from
the second birth to the third. Moreover, results pertaining
to the last transition must be viewed with some skepticism
due to the small number of third births, especially for the
1950 cohort.

Comparisons with Official Statistics. Before turning
to a detailed analysis of births using the 1981-1982 German
Life History Survey, we compare selected summary statistics
based on these data with corresponding estimates based on
official statistics obtained from Birg et al. (1984). Such
comparisons help in assessing the degree to which fertility
of respondents to the 1981-1982 GLHS is representative of
the fertility of the larger West German population. These
comparisons are limited to women because official statistics
for men are not published. Official statistics also do not
provide information on women in the 1930 cohort.

In making these comparisons, two basic differences must
be kept in mind. First, the GLHS sample is limited to
native Germans whereas the official statistics are based on
all permanent German residents. Second, estimates based on
the GLHS sample refer to the first child ever born, ignoring
marital status. The official statistics count the first
birth in the current marriage. Thus, they exclude births
outside marriage, and the first birth within a marriage may
not be the first ever for women who have been married
previously. The estimates based on the official statistics
attempt to correct for the unusual definition of first birth
and the restriction to marriages; still, an estimation
procedure is required. Hence, one should not be too
surprised if the two sets of figures differ somewhat, quite
aside from sampling fluctuations.

Table 3 reports the proportion of women who have not
yet had a first, second, and third child, respectively, by
selected ages for the 1940 and 1950 cohorts as estimated
from the 1981-1982 GLHS and from the official statistics.
All differences are less than .01 for the first birth to
women in the 1940 cohort. Differences for the first birth
to the 1950 cohort are somewhat greater (typically around
.04), but the estimates are not consistently larger or
smaller for one source than for the other. The similarity
of the estimates from the two sources is also clearly
apparent in the case of the second and third births. No

difference exceeds .04, and on average it is much less, only
.016. Overall there are no systematic or large differences
casting doubt on the representativeness of the 1981-1982
GLHS for studying German fertility.

First Birth. Next we examine in more detail the
proportion who are still childless as a function of the
respondent's age in years, sex, and cohort. To those
interested in reasons for the decline in overall fertility
in West Germany, timing of the first birth may seem to be of
little interest because fertility decline is usually
regarded as the result of a sharp decrease in the number of
people having many children. But, postponement of the birth
of the first child can also be a factor in this decline.
First, it has been suggested (Dinkel, 1983; Birg et al.,
1984, Huinink, 1987) that childlessness is becoming more
common among recent German cohorts, and study of the timing
of the first birth can help to assess this. In addition,
postponement of the first birth among those who do
eventually have at least one child shortens the period at
which a person is at risk of having additional children, and
thus can reduce total fertility indirectly.

Panel A of Table 4 shows the first, second and third
quartiles of the age at first birth by sex and cohort.
These estimates come from Kaplan-Meier (1958) estimates of
the survivor functions for the first birth of men and women
in the three cohorts; the point estimates (with 95 percent
point-wise confidence intervals) are displayed graphically
in Figures A.1 through A.3 in the appendix. For women, the
first quartile of age at first birth occurs at the lowest
age for the 1950 cohort and at the highest age for the 1930
cohort. The latter may seem surprising. We think that
hardships in the period after World War II led those in the
1930 cohort to delay family formation. Men in the 1930
cohort also tended to remain childless longer than men in
the 1940 cohort, but not as long as men in the 1950 cohort.

The difference between the first and third quartiles, a
measure of spread in the distribution of age at first birth,

is smallest for the 1940 cohort and largest for the 1950
cohort of women (and, most likely, of men, too).[4] Thus, the
timing of the first birth varies least in the 1940 cohort,
somewhat more in the 1930 cohort, and still more in the 1950
cohort. Estimates based on the official statistics show
that the spread of the age at first birth declined from
older cohorts to the 1946 cohort, when variation in this
timing reached a miminum, and then began to increase again.
Thus, external evidence suggests that this pattern is not a
peculiarity of the GLHS sample.

It is easy to understand why those in the 1930 cohort,
who postponed their first birth during a period of national
recovery, tended to have their first child in a relatively
narrow age range. It is less clear why age at first birth
varies relatively little in the 1940 cohort. Most members
of this cohort had their first child during the early 1960s,
a period of economic expansion, an improving standard of
living, and a relaxation of traditional norms of sexual
behavior. The latter may have encouraged members of the
1940 cohort to have children at an earlier age. Moreover,
the favorable economic conditions may have tended to reduce
variation in timing of the first child, just as unfavorable
conditions in the post-war period did for the 1930 cohort.
That is, very favorable as well as very unfavorable societal
conditions may dampen variation in the timing of births that
occurs due to heterogeneity in personal attributes. Hence,
the greater variation in the timing of the first birth in
the 1950 cohort suggests that individual characteristics may
explain entry into parenthood for the 1950 cohort more than
for the others. Subsequent analyses of birth rates have
borne this out, as we report in our companion paper.

Another change over time in fertility patterns that can
be seen in Panel A of Table 4 (and even more clearly in
Figures A.1 through A.3 in the appendix) pertains to the
differences between men and women in the timing of the first
birth. As expected, for each cohort men tend to have their
first child at an older age than women. In addition, the

difference between men and women in age at first child has
increased over time. For each quartile of the distribution
of the age at first birth, the male-female difference is
least for the 1930 cohort (where it ranges from 2.3 to 3.2
years) and greatest for the 1950 cohort (where it ranges
from 4.7 to 5.1 years).

This finding is somewhat surprising since the fraction
of married women employed began to rise slowly but steadily
after 1967-1968 (see Figure 1) and because the ideology of
women's liberation (i.e., sex role equality) received much
attention in West Germany during the 1970s when the 1950
cohort was, for the most part, having a first child. Yet,
this period emphasizing women's economic contributions and
equality between the sexes reveals increased _inequality_ in a
major sex-linked behavior---birth of the first child.
Moreover, since the ideology of sex role equality has tended
to encourage women's activities outside the home more than
men's activities in the home (Davis, 1984), one might have
expected this ideology to affect women's fertility behavior
more than men's. But the opposite appears to be the case.

The large difference between men and women in the 1950
cohort in the timing of the first birth may be explained by
the birth year of the spouse. On average German women marry
men roughly three years older than themselves (see Table 1).
Hence women are mainly married to men from older cohorts,
and men are mainly married to women from younger cohorts.
Since the ideology of sex role equality has been accepted
more widely among younger Germans, the spouses of the men in
the 1950 cohort are more likely to work than the women in
this cohort. So, it is perhaps not so surprising after all
that fertility of men in the 1950 cohort appears more
consistent with an impact of the new ideas about women's
roles than does the fertility of the women in this cohort.

Another way to study differences between men and women
in different cohorts with respect to the birth of the first
child is to compare their age-specific birth rates.[5] Plots
of the Nelson-Aalen[6] estimates of the rate of first birth

(smoothed over a year) versus age are displayed by cohort for men and women in Figures 2 and 3, respectively.

Figure 3 suggests that the first-birth rates of women in the 1930 and 1940 cohorts were fairly similar under 22 years. The main difference between these two groups is that the first-birth rate for women between 22 and 25 years is markedly higher in the 1940 cohort than in the 1930 cohort. Women in the 1930 cohort eventually "catch up" by a higher first-birth rate between ages 25 and 35 years than women in the 1940 cohort.

The first-birth rate of women in the 1950 cohort is noticeably different from that of women in the other two cohorts. There is a marked bimodality in the distribution of the first-birth rate, with one peak occurring about age 21 years and another about 26 years. A similar pattern has also been observed in the official statistics for slightly younger cohorts, so it is unlikely to be a peculiarity of the GLHS sample.

There are at least two possible explanations for this bimodality. First, it may result from observing a mixture of two (or more) groups with rather strikingly different norms concerning when to begin family formation. According to this explanation, one group is similar to the 1940 cohort of women and tends to begin family formation early; another group postpones the birth of the first child. The first group may adhere to more traditional values about women's roles whereas the second may be more accepting of modern values and more interested in women's enhanced opportunities for employment. Variables that might differentiate fairly well between these two postulated groups are education and training. We expect women to fall largely in the first group if they had relatively little schooling and training, and in the second group if they had relatively a lot.

Second, the bimodality may reflect a pure "period" effect. The marked drop in the birth rates between ages 22 and 25 for the 1950 cohort of women occurs around 1972-1975. In 1973-1974, oil prices rose sharply in Western Europe,

causing a period of economic uncertainty and decline relative to the boom in the late 1960s. These economic and social changes in German society may have caused many young women in the 1950 cohort to postpone family formation for a few years. Figure 4 displays indicators of economic activity in the Federal Republic of Germany over time (gross national product, i.e., GNP, in Panel A; percentage change in GNP in Panel B) to give some idea of economic conditions during the period when members of this cohort were likely to have their first child.

To try to assess the first explanation of the bimodal pattern, we estimated the first-birth rate of women in the 1950 cohort whose highest educational degree did and did not exceed the "Hauptschule Abschluss," the degree from an elementary school, which corresponds to about nine years of completed schooling (see Figure 5). Bimodality is clearly apparent for the less educated group of women, but not for the more educated women, for whom the rate of first birth has a single peak in the late twenties. Next we estimated the first-birth rate of less educated women with and without a vocational training degree (see Figure 6). Although bimodality is still apparent, it is less than in Figure 5. Moreover, since the primary peak for those without training is at the younger age whereas the primary peak for those with training is at the older age, this picture is consistent with the hypothesis that there are two groups, one oriented to early family formation and another to working for a few years before having a first child. But there may also be a behavioral response to the unfavorable economic conditions in 1973-1975. These analyses do not provide evidence on this other possibility.

The patterns of the age-specific rate of first birth of men (see Figure 2) are fairly similar to those for women, though differences between the cohorts tend to be smaller. Men in the 1930 cohort tend to postpone births in their late twenties and early thirties, but eventually "catch up" with men in the 1940 cohort through higher first-birth rates at

ages 33 through 38 years. The first-birth rate of men in
the 1950 cohort is not bimodal. The most striking pattern
for this cohort of men is just their very low first-birth
rate at all ages yet observed. One cannot help but wonder
just what fraction will ultimately remain childless.

With regard to this, consider the proportion childless
as a function of age (see Panel A of Table 5). It is
striking that nearly half of the men in the 1950 cohort are
still childless by age 30. Contrast this figure with the
figures for men in the 1930 and 1940 cohorts, .38 and .30,
respectively, and also with those for women in the 1930,
1940, and 1950 cohorts, .24, .18, and .27, respectively.
Although men are biologically capable of bearing children
for far more years than are women, Figures 2 and 3 indicate
that, for the older cohorts, men over 35 years of age have
only slightly higher birth rates than women, and that birth
rates are very low for both sexes after this age. This
suggests that men in the 1950 cohort will either break this
pattern and have much higher birth rates in their thirties
(and perhaps at older ages) than men in previous cohorts, or
remain childless throughout their lives to a truly amazing
degree. Since the proportion of childless women in the 1950
cohort at any given age does not differ from that of women
in the older cohorts to such a considerable extent, these
men may largely be postponing child-bearing rather than
permanently avoiding fatherhood. On the other hand, since
men tend to marry women about three years younger, their
high degree of childlessness may reflect the greater
tendency of women in still younger cohorts to engage in
market work rather than house work and to choose careers
over children. New data are needed to answer this question.

With the GLHS data, the proportion ultimately childless
can be estimated only for the 1930 and 1940 cohorts. An
indicator of this quantity is the proportion childless at
age 40 (see Panel A of Table 5). Since the confidence
interval at this age is fairly wide, one must be cautious
about comparing the point estimates. Still, the proportion

childless at age 40 is remarkably similar for the 1930 and 1940 cohorts for a given sex, though about .03-.04 more men than women seem to be childless at this age. One forecast predicts that .18 of the 1950 cohort of women will be childless ulitmately (Huinink, 1987). This could be too high since the proportion of childless women in the 1930 cohort dropped from .24 at age 30 to .11 at age 40. Note that the proportion of childless women at age 30 is .27 for the 1950 cohort, which is only .03 higher than for the 1930 cohort. Only new data can show to what extent it is becoming more common to remain childless over the life span.

Second Birth. The Kaplan-Meier estimated probability that men and women have not yet had a second child is given as a function of age for the 1930, 1940 and 1950 cohorts in Panel B of Table 5 and displayed graphically in Figures A.4 through A.6 in the appendix. Several patterns noted in the case of the first birth are again apparent, but not all. Again the 1940 cohort is noteworthy for tending to have a second child at a younger age and in a smaller range of ages than the other two cohorts, especially the 1950 cohort. As in the case of the first child, having a second child tends to occur several years later for men than women, and the male-female difference in age at the child's birth increases across the cohorts. Finally, the relatively small fraction of men and women having a second child between ages 35 and 40 reveals a marked decline in the birth rate after age 35.

The proportion who ultimately have at least two children can be approximated from the proportion at age 40 for the 1930 and 1940 cohorts. For women, this figure is .676 and .659, respectively; for men, .628 and .611. The intercohort differences for individuals of the same sex are too small to be taken seriously given the width of the 95% confidence intervals. The male-female difference in a given cohort is probably genuine, but it seems to reflect mainly men's late start in child-bearing rather than ultimate differences in achieved parenthood. In sum, for the two older cohorts, roughly one-third have one or fewer children.

Since we reported above that about 11 percent in these cohorts remain childless, about 22 percent (= 33% - 11%) have exactly one child. In contrast, about 15 percent of the respondents come from one-child families.

Since members of the 1950 cohort were in their early thirties at the interview, one cannot accurately estimate the fraction that will ultimately have more than one child. Note, however, that the proportion of 30-year old women with at least two children is very similar for the 1930 and 1950 cohorts, .452 and .438, respectively. The proportion of 30-year old men with at least two children is much smaller (.228) in the 1950 cohort, but it does not differ strikingly from that for the other two cohorts of men, in contrast to the case of the first birth. Since men in the 1950 cohort are childless to a much greater extent than men in the other cohorts at any age yet observed, a similar proportion with two children implies that most men in the 1950 cohort who have one child go on to have two. This points to another form of heterogeneity in the 1950 cohort that is similar to that for women: one group remains childless (or delays child-bearing until ages beyond those observed) and a second group has rather traditional fertility patterns.

Third Birth. As we noted earlier, the occurrence of a third birth is relatively uncommon in these data, especially in the 1950 cohort. Consequently, one cannot say much at all about the third birth for this cohort, and only a little about it for the other two cohorts.

Panel C of Table 5 gives the Kaplan-Meier estimated probability of not yet having a third child by age, sex, and cohort. The greatest difference between the 1930 and 1940 cohorts is in the proportion having a third child by age 40, which falls from .300 to .179 in the case of men, and from .324 to .273 in the case of women. Recall that members of the 1930 cohort tended to have their first and second children at older ages than members of the 1940 cohort. Yet by age 40, the overall fertility of those in the 1930 cohort has not only caught up with that of the 1940 cohort, but

exceeded it. This finding illustrates that <u>postponement</u> of the first birth need not lower <u>completed</u> fertility.

<u>Birth Spacing</u>. The probability of having a second or third child by a given age depends on the timing of previous births. For example, the rise in the male-female difference in the age at the second birth across the cohorts may result primarily from the increasing male-female difference in the age at first birth. An examination of the <u>spacing</u> of births after the first helps to clarify such issues.

Panel A of Table 6 reports the Kaplan-Meier estimate of the probability of not yet having a second child for selected durations since the first child by sex and cohort. The differences between men and women in a given cohort are very small, as are the differences between the cohorts. The overall impression is one of much greater similarity than in the timing of the first birth. Women in the 1950 cohort tend to delay the birth of the second child slightly longer than women in the other two cohorts, but this is the only noticeable difference among the groups.

Panel B of Table 4 gives the first, second, and third quartiles of the duration from the first to the second birth. Twenty-five percent have a second birth within a little over two years after the first birth, and 50 percent have a second birth within 3.9-4.5 years. Again, there are no especially noteworthy differences between men and women or between the different cohorts in this pattern.

Figures 7 and 8 show plots of Nelson-Aalen estimates of the rate of transition from birth one to birth two as a function of the duration since birth one by cohort for men and women, respectively. For both sexes in all three cohorts, the rate is identically zero under a half year, and it is also very low in the second half year, undoubtedly for biological reasons. Between one and five years after the first birth, the second-birth rate tends to be relatively high for all groups, although its shape and timing of the peak varies. Women (and, to a lesser extent, men) in the 1950 cohort have a noticeably lower second-birth rate

between 1.0-1.5 years after the first child than do the
other groups; probably this results from the widespread
adoption of effective contraceptive methods by married
couples in the 1970s in order to space births as well as to
avoid unwanted births. Indeed, a spacing of about three
years appears especially popular. Between five and ten
years after the first birth, the second-birth rate is
declining but not negligible. In fact, the Kaplan-Meier
estimates of the proportion without a second child falls on
average by about .15 between 5 and 10 years after the first
child's birth (see Panel A of Table 6), which indicates that
the spacing between the first and second births is quite
long in a substantial fraction of cases.

   As we mentioned before, due to the rarity of third
births in the data, one can only gain a general impression
of the transition from the second to the third birth. Panel
B of Table 6 gives the Kaplan-Meier estimated probability of
not yet having a third child among those with a second child
for selected durations by sex and cohort. Male-female
differences are small; cohort differences are larger. As
one expects, the more recent the cohort, the smaller the
proportion who have had a third birth after any given
duration since the second birth. The greater frequency of
at least three children among the 1930 cohort than among the
other cohorts is especially apparent.

   Nelson-Aalen estimates of the rate of transition from
the second to the third birth are shown in Figures 9 and 10
by cohort for men and women, respectively. For any given
duration since the second birth, these rates tend to be much
lower than the rate of transition from the first to the
second birth (cf. Figures 7 and 8). Moreover, they are much
less peaked. The low level and flatness of the third-birth
rate for the 1950 cohort of women is especially evident,
suggesting that third births are often unintended and
unplanned in this cohort.

   Variation across the cohorts seems quite clear for men:
in the first three years after the second birth, which is

when the majority of third children are born, the third-
birth rate tends to fall from the 1930 to 1940 to 1950
cohort. The pattern for women is similar, although the
differences between the 1930 and 1940 cohorts are less
marked than for men.

Summary. First, completed family size falls as one
moves from the 1930 to the 1950 cohort. This appears to be
mainly due to a significant decrease in the fraction having
at least three children. There is no strong suggestion that
childlessness is becoming much more common, except possibly
in the 1950 cohort of men. But, they are in their early
thirties at the interview and have many remaining years in
which they could bear children.

Second, the rate of transition from the second birth to
the third is quite low and relatively flat for both men and
women in all three cohorts. The low level suggests that a
third child is rarely intended; the flatness suggests that
the third child is usually not planned.

Third, spacing between the first and second births is
remarkably similar for men and women in all cohorts. The
only deviation from this pattern of overall similarity is a
reduction in the likelihood of a very short spacing (i.e.,
under 18 months) in the 1950 cohort, which suggests that
members of this cohort use contraceptives for birth spacing
as well as birth avoidance. In all three cohorts a spacing
of about three years is especially common.

Fourth, the tendency to postpone the birth of the first
child is especially noticeable for men relative to women,
and for the 1930 and 1950 cohorts relative to the 1940
cohort. The difference between men and women presumably
reflects traditional norms about mate selection. We
suggested that the 1930 cohort tended to postpone births
during the post-war period, when they were of an age at
which they might normally have been expected to have a first
child, while recovering from the hardships resulting from
World War II. We think that the 1950 cohort tended to
postpone the first birth for very different reasons, in

particular, because of the increasing opportunities for women to engage in market work and perhaps also because of economic uncertainty and decline in the mid-1970s.

Fifth, even though members of the 1930 cohort tended to postpone their first birth, they also exhibit a marked tendency to "catch up" at older ages. Fertility of this cohort shows clearly that postponement of the first birth need not mean permanent avoidance of parenthood.

Sixth, variation in the timing of the first birth is greatest in the 1950 cohort. Thus, there seems to be more variation in fertility to be explained by individual differences for this cohort than for the older ones. We suggested that there may be relatively little variation in this timing for the 1930 cohort because most delayed child-bearing during the post-war recovery and for the 1940 cohort because relaxation of sexual norms and favorable economic conditions encouraged most people to begin child-bearing early. Thus, either very favorable or very unfavorable general social and economic conditions can produce a social situation in which individual attributes are relatively unimportant in determining fertility.

Seventh, the rate of first birth of women in the 1950 cohort as a function of age is clearly bimodal. This seems to be partly due to heterogeneity across social groups in the fundamental shape of the birth rate pattern. It may also be partly a period effect due to the adverse economic conditions in 1973-1974. We suggest that the heterogeneity may result from a mixture of two groups, one whose views of women's roles are more traditional and chose to have children at relatively young ages, and another whose views are more modern and oriented to women working outside the home, which leads to postponement of the first child. We found that level of completed education and whether a woman has a vocational training degree discriminate among fertility patterns and substantially reduce the extent of bimodality in the first-birth rate of women in this cohort.

ENDNOTES

[1]The latter figure is even more suprising than that for native Germans because it is so much lower than for these populations in their native lands. Fertility is also low in the German Democratic Republic, though not to this degree.

[2]A very small number of respondents were interviewed in early 1983. Designating this the "1981-1982" study gives the best picture of its timing for the majority of the respondents.

[3]The respondents were selected through a sampling plan based on the "ADM" sampling design, which involves three steps: (1) a stratified sampling of voting districts in the Federal Republic of Germany, in which the probability of selection is proportional to the number of households in the district, (2) random sampling of households in the voting districts chosen, and (3) random sampling of exactly one respondent per household who was born in the desired years from all households that have at least one member in the desired cohorts. For further information, see Hoffmeyer-Zlotnik et al. (1984).

[4]By age 32.4 years, .37 of the men in the 1950 cohort are still childless. Unless .12 of the men in this cohort had their first child between ages 32.4 and 33.4 (which seems extremely improbable), the interquartile range for men will turn out to be largest for those in the 1950 cohort.

[5]Examination of log survivor plots is often recommended to see whether rates vary over time. Here we have strong reason to expect age variation a priori, so this step is not very informative. In any case, plots of the rates versus age are easier to decipher.

[6]This estimator was first proposed by Nelson (1972); later Aalen (1978) gave it a firmer statistical foundation. It has therefore come to be called by the names of both men.

REFERENCES

Aalen, O.O.   (1978)

   "Nonparametric Inference for a Family of Counting

   Processes," Annals of Statistics 6(4): 701-726.

Birg, H., J. Huinink, H. Koch, and H. Vorholt.   (1984)

   Kohortenanalytische Darstellung der Geburtenentwicklung

   in der Bundesrepublik Deutschland.   IBS Materialen Heft

   10.   Bielefeld.

Davis, K.   (1984)

   "Wives and Work: The Sex Role Revolution and its

   Consequences," Population and Development Review 10(3):

   397-417.

Dinkel, R.   (1983)

   "Analyse und Prognose der Fruchtbarkeit am Beispiel der

   Bundesrepublik Deutschland," Zeitschrift fuer

   Bevoelkerungswissenschaft 9(1): 47-72.

Hoffmeyer-Zlotnik, J., H.P. Kirschner, M. Wiedenbeck, and

   E. Brueckner.   (1984)

   "Methodenreport ´Lebensverlaeufe´", ZUMA-Technischer

   Bericht Nr. T84108.   Mannheim.

Huinink, J.   (1987)

   "Kohortenanalyse der Geburtenentwicklung in der

   Bundesrepublik Deutschland," in A. Herlth and

   K.P. Strohmeier (eds.), Lebenslaeufe und

   Familienentwicklung. Leverkusen, forthcoming.

Kaplan, E.L., and P. Meier.   (1958)

   "Nonparametric Estimation from Incomplete

   Observations," Journal of the American Statistical

   Association 53(282): 457-481.

Kiefl, H., and J. Schmid. (1985)
    Empirische Studien zum generativen Verhalten. Boppard.

Michael, R.T., and N.B. Tuma. (1985)
    "Entry into Marriage and Parenthood by Young Men and
    Women: The Influence of Family Background," Demography
    23(4): 515-543.

Nelson, W. (1972)
    "Theory and Applications of Hazard Plotting for
    Censored Failure Data," Technometrics 14(4): 945-966.

Strohmeier, K.P. (1985)
    Familienentwicklung in Nordrhein-Westfalen.
    Schriftenreihe des Ministerpraesidenten des Landes
    Nordrhein-Westfalen Heft 47. Duesseldorf.

Table 1. Means of Selected Covariates by Sex and Cohort

	Men			Women		
	1930	1940	1950	1930	1940	1950
D R's Educ.: no "Abschluss"	.116	.079	.059	.102	.063	.059
D R's Educ.: "Abschluss"	.686	.703	.525	.699	.659	.597
D R's Educ.: "mittlere Reife"	.116	.130	.194	.142	.194	.213
D R's Educ.: "Abitur"	.082	.088	.222	.057	.084	.131
D R had Vocational Training	.604	.742	.689	.280	.457	.588
N of Children of R's Parents	3.62	3.27	3.01	3.70	3.32	3.16
N of R's Own Children	2.13	1.76	0.92	2.17	1.98	1.39
D Catholic	.527	.445	.453	.488	.439	.473
D Large Place when R Age 14	.177	.264	.291	.151	.248	.249
D Small Place when R Age 14	.427	.439	.356	.512	.409	.350
D One Parent when R Age 14	.104	.179	.074	.087	.224	.067
Age Difference of Spouse	-2.83	-2.49	-2.83	3.21	3.23	3.14
Number of Cases	347	375	365	361	355	368

Note on definitions of variables: A "D" before a variable indicates a dummy variable in which 1 = the category named. With regard to educational level, "[Hauptschule] Abschluss" means elementary school with a degree (about nine years), no "Abschluss" means some elementary school but without a degree, "mittlere Reife" is roughly equivalent to some high school education (approximately 10 years), "Abitur" means graduation from a gymnasium (at least 13 years). A "large" place is defined as one with more than 100,000 residents, and a "small" place is one with fewer than 20,000 residents.

Table 2.   Counts of Own Children by Sex and Cohort

Panel A.   Exact Count

Number of Births	Men			Women		
	1930	1940	1950	1930	1940	1950
0	45	58	154	38	39	86
1	79	85	101	77	82	106
2	110	175	94	126	137	136
3	64	42	16	60	62	30
4	28	15	0	38	23	6
5	8	4	0	10	8	3
6	9	3	0	3	2	0
7	3	3	0	5	0	1
8	3	0	0	2	2	0
Mean	2.13	1.76	0.92	2.17	1.98	1.39

Panel B.   Cumulative Count

Number of Births	Men			Women		
	1930	1940	1950	1930	1940	1950
1	304	317	211	321	316	282
2	225	232	110	244	234	176
3	115	67	16	118	97	40
4	51	25	0	58	35	10
5	23	10	0	20	12	4
6	15	6	0	10	4	1
7	6	3	0	7	2	1
8	3	0	0	2	2	0

Table 3. Proportion of Women with Fewer than N Births
by Age and Cohort

		N = 1		N = 2		N = 3	
		1940	1950	1940	1950	1940	1950
Age 20	GLHS	.836	.791	.966	.981	.997	1.000
	Official	.840	.758	.975	.961	.998	.997
Age 25	GLHS	.381	.473	.731	.791	.929	.970
	Official	.374	.434	.749	.789	.936	.962
Age 30	GLHS	.172	.253	.441	.552	.822	.890
	Official	.181	.293	.484	.589	.821	.904
Age 35	GLHS	.118	-	.385	-	.751	-
	Official	.118	-	.391	-	.762	-

Table 4. First, Second, and Third Quartiles of the Timing
of Selected Birth Intervals by Sex and Cohort

Panel A. Age (in years) at First Birth

	Men			Women		
	1930	1940	1950	1930	1940	1950
25%	24.8	24.4	25.5	22.3	21.5	20.8
50%	28.1	27.2	29.9	25.8	23.8	24.8
75%	32.6	31.5	-	29.4	27.8	30.8

Panel B. Duration (in years) between First and Second Births

	Men			Women		
	1930	1940	1950	1930	1940	1950
25%	2.2	2.2	2.3	2.0	2.2	2.6
50%	4.3	3.9	4.2	4.1	4.1	4.5
75%	16.8	12.0	-	13.2	15.2	10.6

Table 5.   Proportion with Fewer than N Children
            by Age (in years), Sex, and Cohort

			Age		
	20	25	30	35	40
Panel A. N = 1					
Men					
1930	.968	.739	.384	.186	.132
1940	.984	.699	.301	.181	.155
1950	.984	.792	.496	–	–
Women					
1930	.910	.551	.236	.116	.105
1940	.864	.402	.180	.121	.110
1950	.820	.489	.266	–	–
Panel B. N = 2					
Men					
1930	.997	.962	.703	.455	.372
1940	.997	.920	.640	.445	.389
1950	.997	.958	.772	–	–
Women					
1930	.994	.850	.548	.377	.324
1940	.986	.780	.467	.369	.341
1950	.989	.809	.562	–	–
Panel C. N = 3					
Men					
1930	.997	.988	.914	.778	.700
1940	.997	.987	.925	.861	.821
1950	.997	.994	.978	–	–
Women					
1930	.997	.931	.875	.726	.676
1940	.997	.941	.828	.755	.727
1950	.997	.984	.905	–	–

Table 6. Proportion with Fewer Than N Children by Duration
since Previous Birth (in years), Sex, and Cohort

	Duration			
	2	5	10	15
Panel A. N = 2				
Men				
1930	.773	.444	.298	.274
1940	.797	.380	.272	-
1950	.799	.424	.262	-
Women				
1930	.760	.411	.274	.246
1940	.781	.402	.270	.253
1950	.831	.469	.259	-
Panel B. N = 3				
Men				
1930	.862	.622	.496	-
1940	.882	.788	.705	-
1950	.931	.744	-	-
Women				
1930	.848	.663	.540	.519
1940	.858	.685	.607	-
1950	.910	.752	.606	-

Note: A dash indicates no information.

Figure 1

PERCENTAGE OF MARRIED WOMEN EMPLOYED BY AGE AND YEAR

Percent Employed

100
90
80
70
60
50
40
30
20
10
0

1960   1965   1970   1975   1980

Year

LEGEND

Age 20-25
Age 25-30
Age 30-35
Age 35-40
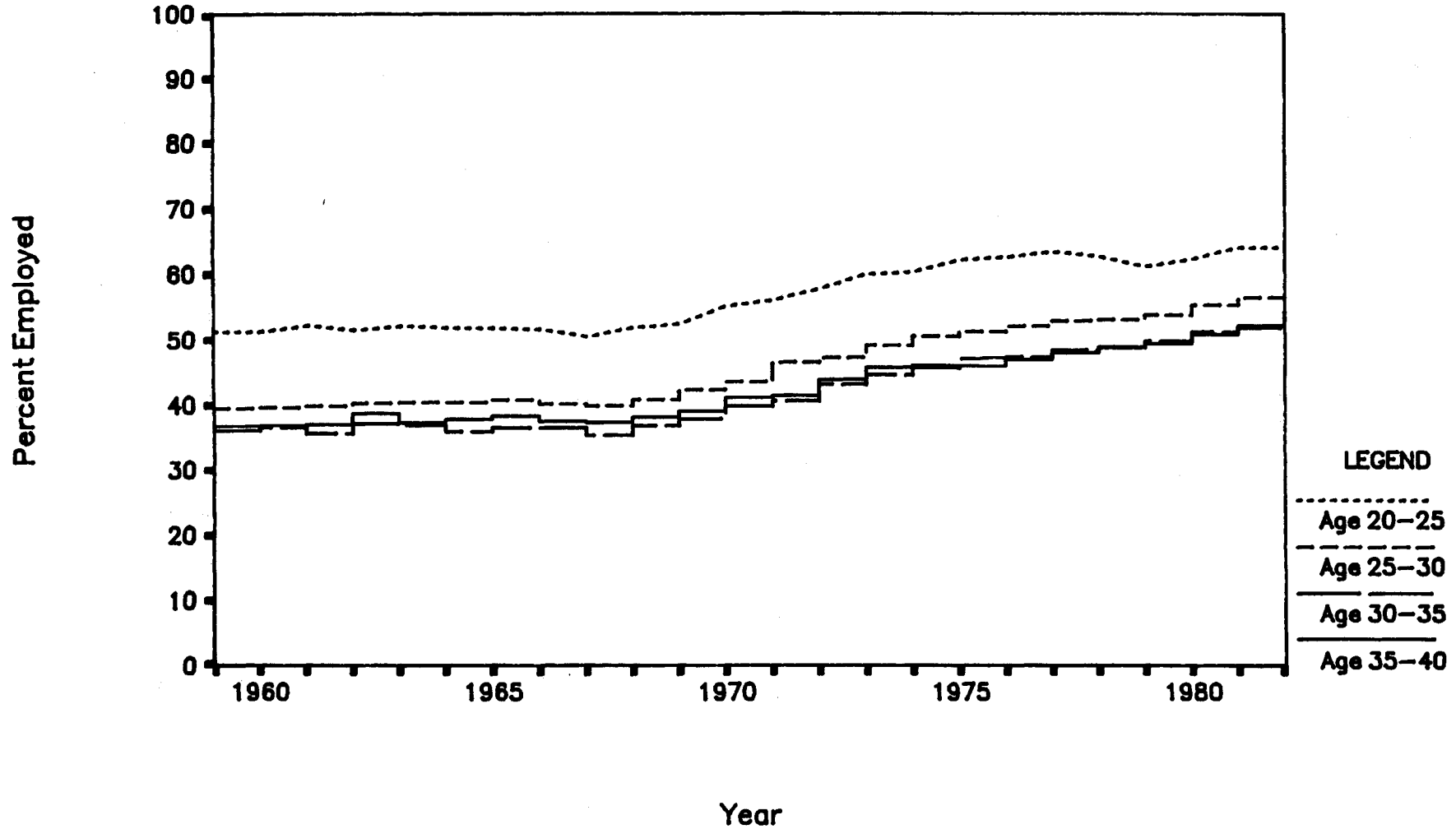
Source : Statistical Yearbook of the Federal Republic of Germany

Figure 2

FIRST BIRTH RATE VS AGE
MEN BY COHORT

LEGEND

.... 1950 Cohort

— — — 1940 Cohort

———— 1930 Cohort

First Birth Rate

Age in Years

1981—82 German Life History Survey

FIRST BIRTH RATE VS AGE
WOMEN BY COHORT

Figure 3

LEGEND

.... 1950 Cohort

– – – – 1940 Cohort

———— 1930 Cohort

First Birth Rate

Age in Years
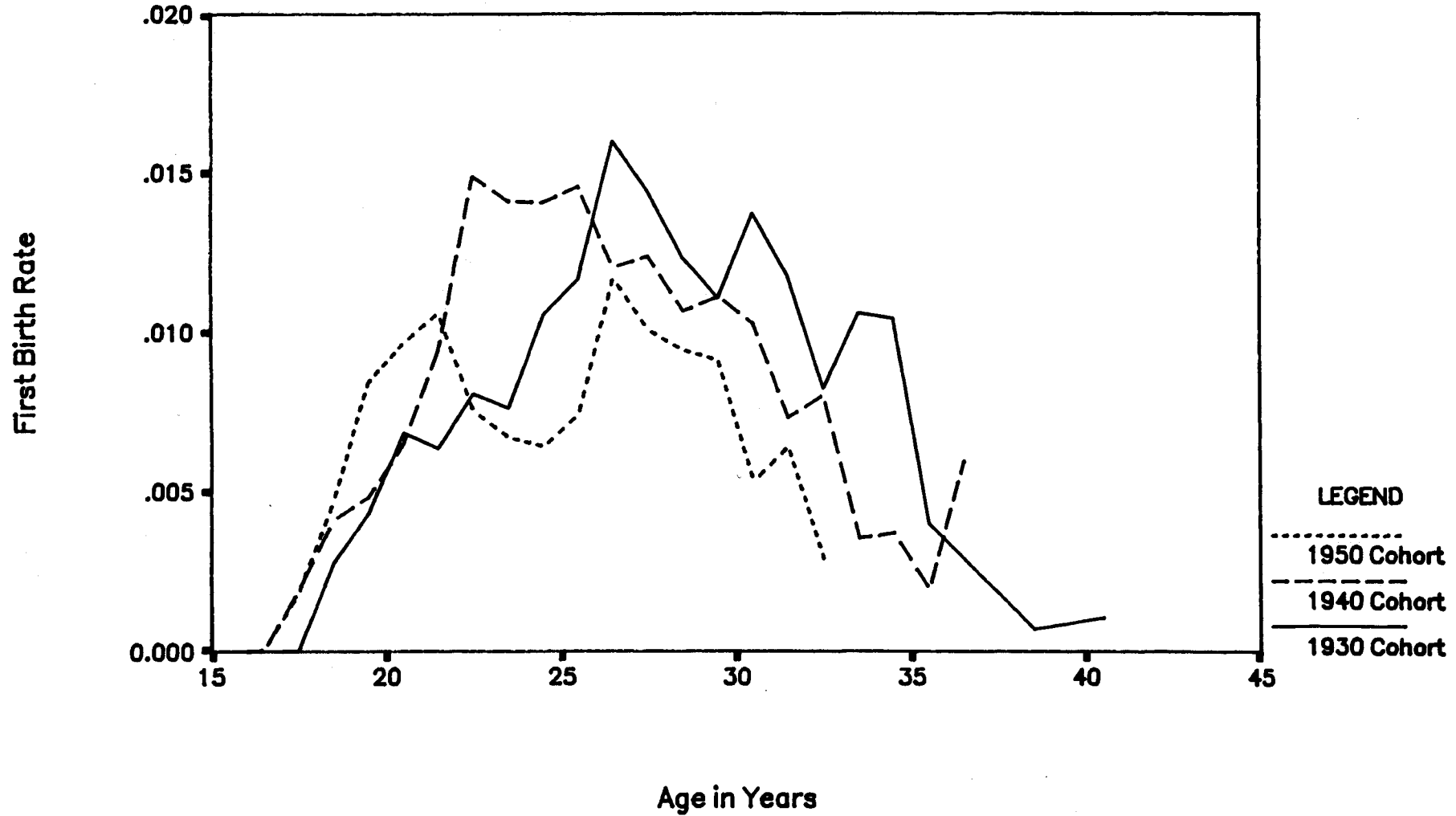
1981—82 German Life History Survey

Figure 4a

GROSS NATIONAL PRODUCT 1960 – 1982
IN 1976 PRICES
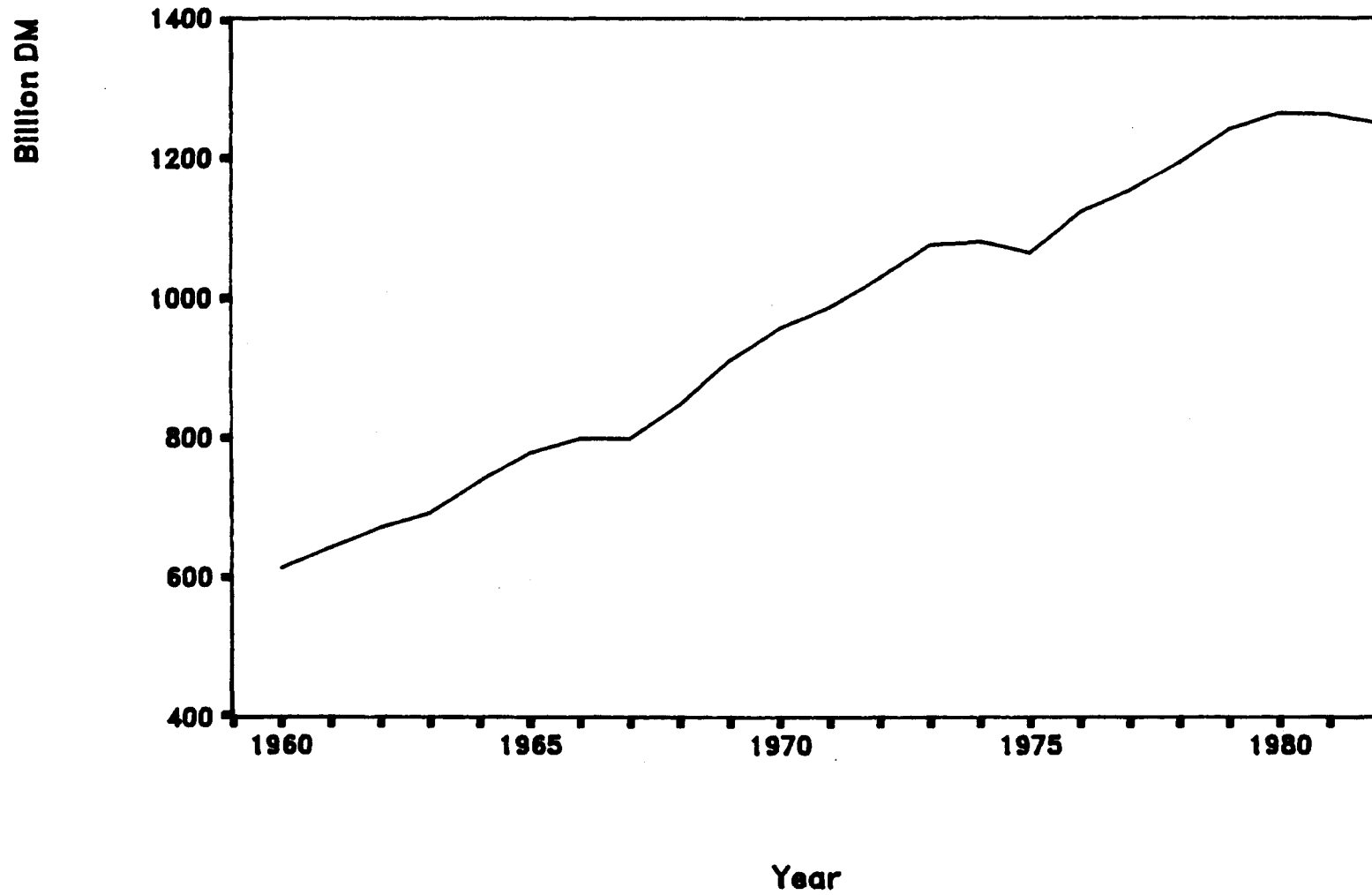
Year

Source : Statistical Yearbook of the Federal Republic of Germany

Figure 4b

GROSS NATIONAL PRODUCT 1960 – 1982
RATES OF CHANGE

Rates of Change (In %)
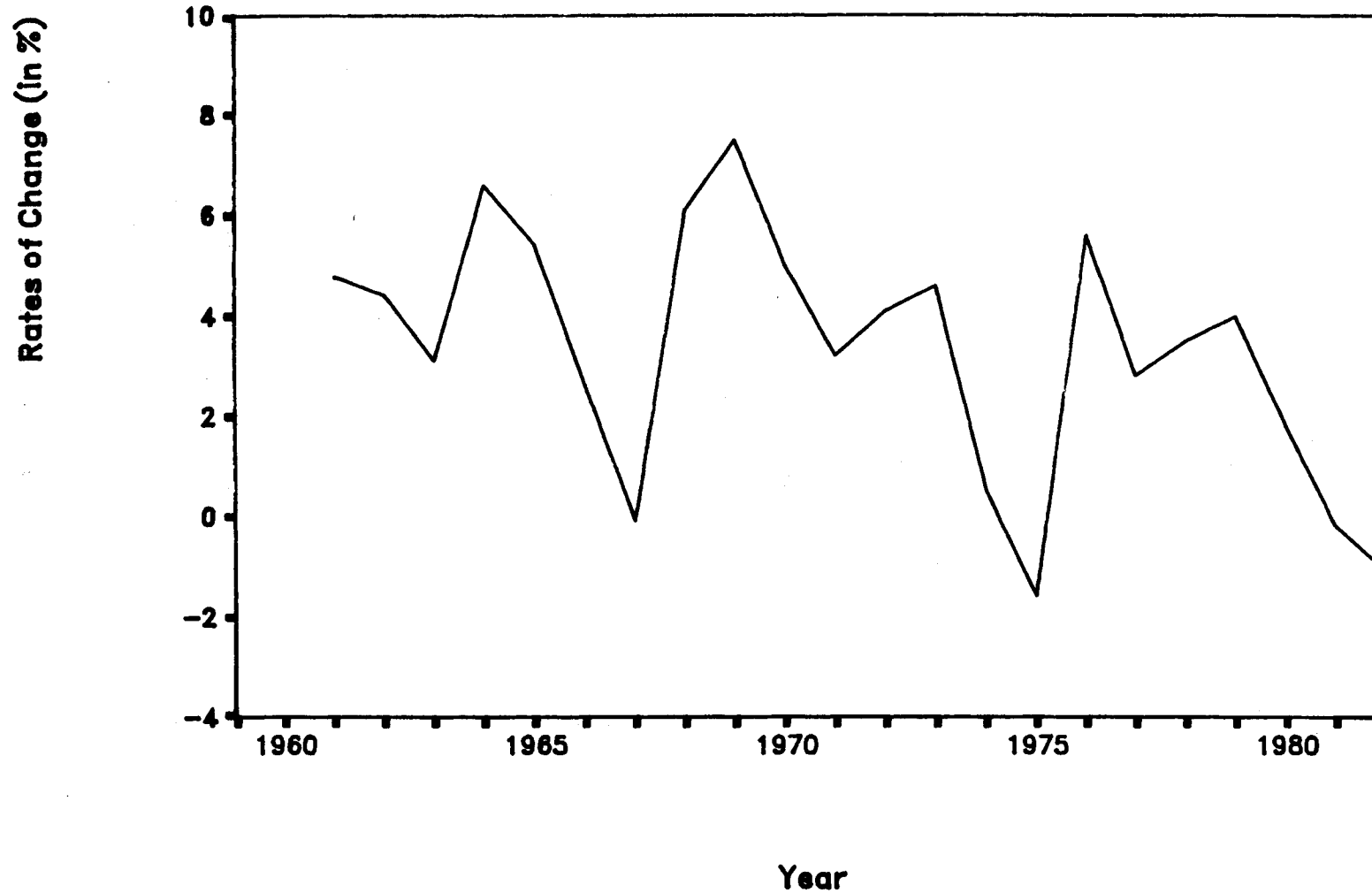
Year

Source : Statistical Yearbook of the Federal Republic of Germany

Figure 5

FIRST BIRTH RATE OF WOMEN IN 1950 COHORT
BY EDUCATIONAL LEVEL

Age in Years

First Birth Rate

LEGEND

Educ levels 1 & 2

Educ levels 3 & 4
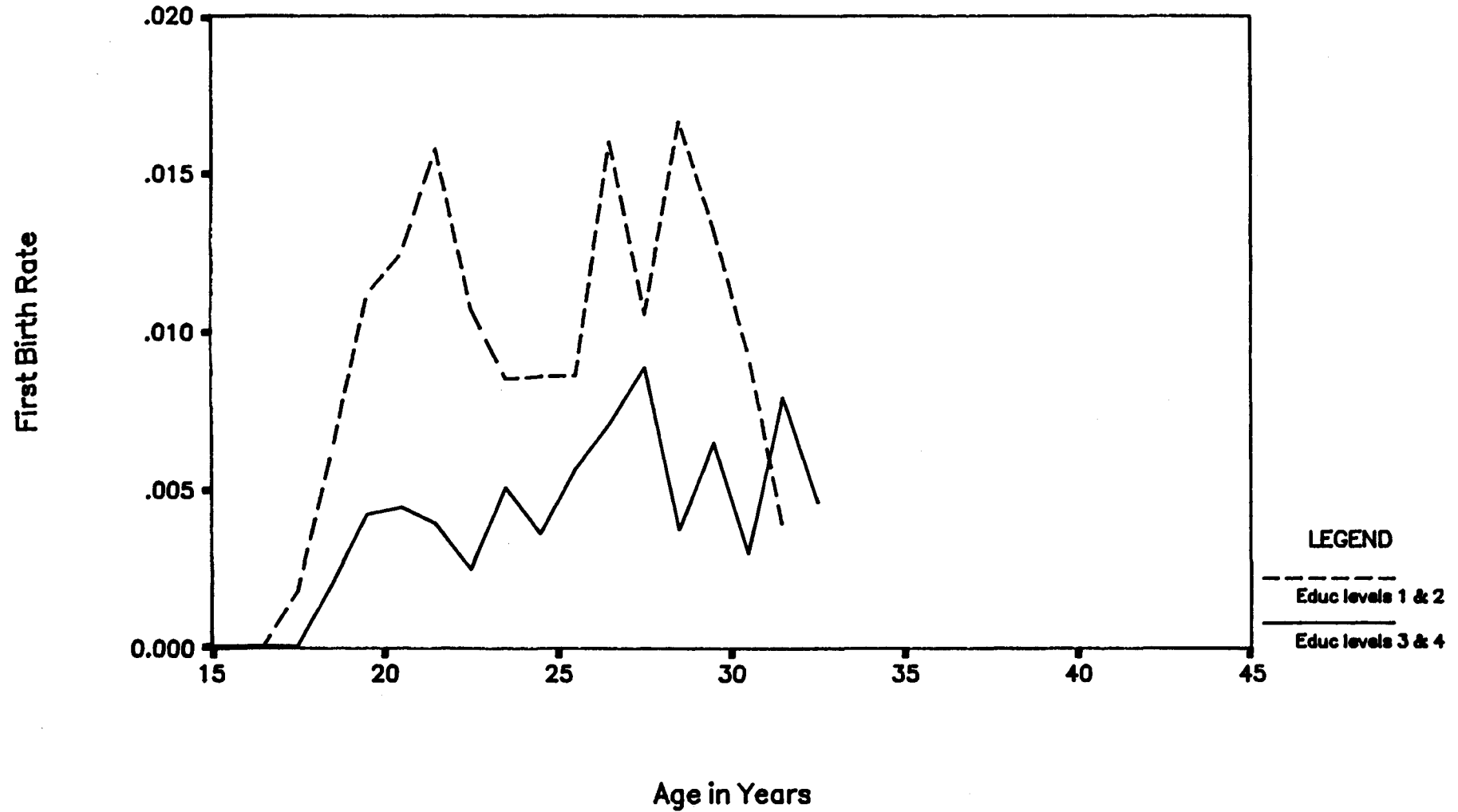
1981—82 German Life History Survey

FIRST BIRTH RATE OF WOMEN IN 1950 COHORT WITH EDUCATIONAL LEVELS 1 OR 2 BY TRAINING

Figure 6

LEGEND

No Training

Training

First Birth Rate

Age in Years

1981—82 German Life History Survey

Figure 7 SECOND BIRTH RATE VS DURATION SINCE FIRST BIRTH
MEN BY COHORT

Second Birth Rate, Given Birth One

Years Since First Birth

LEGEND

1950 Cohort

1940 Cohort

1930 Cohort

1981—82 German Life History Survey

Figure 8 **SECOND BIRTH RATE VS DURATION SINCE FIRST BIRTH WOMEN BY COHORT**

Years Since First Birth
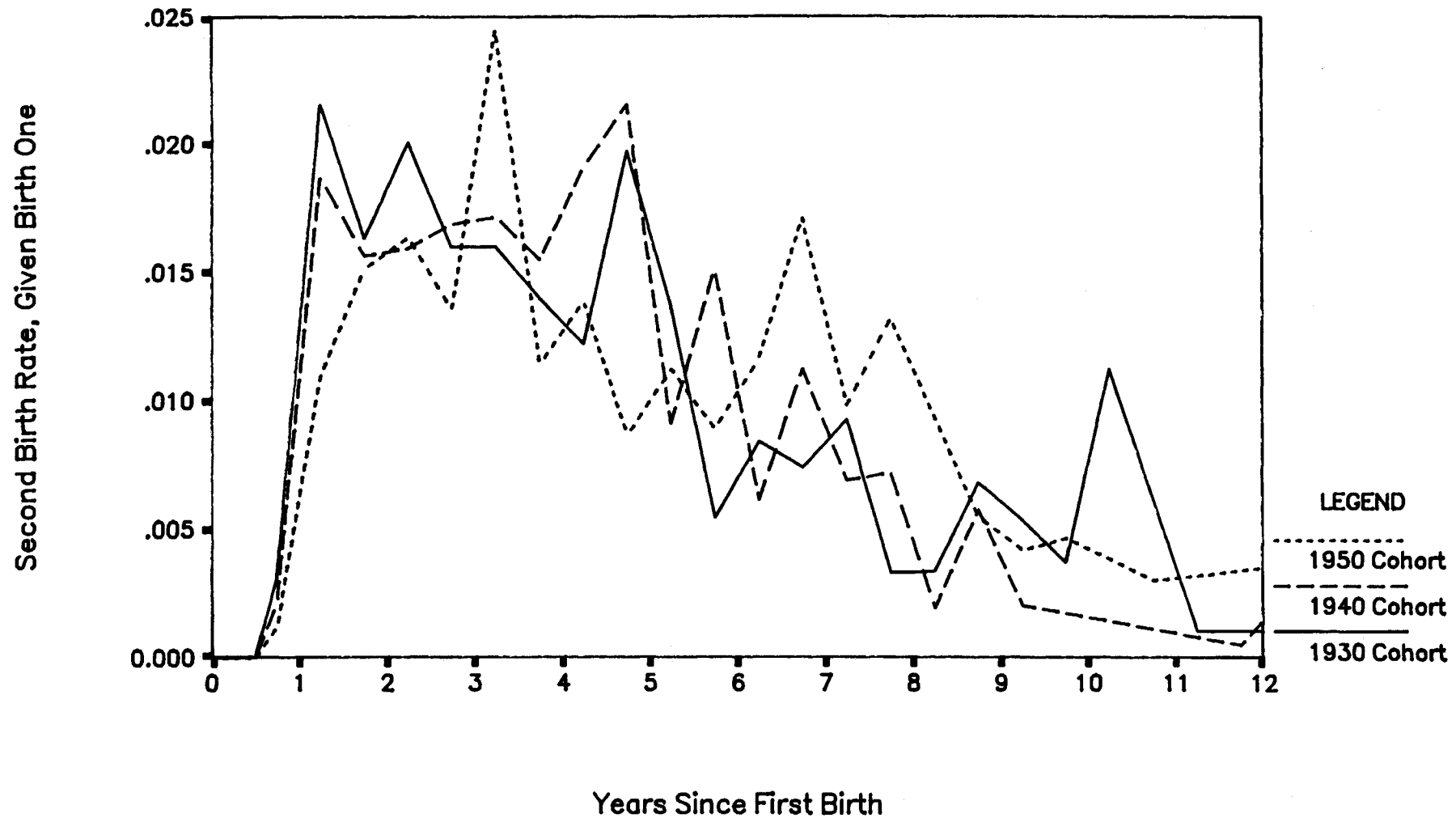
1981—82 German Life History Survey

Figure 9   THIRD BIRTH RATE VS DURATION SINCE SECOND BIRTH
MEN BY COHORT

Third Birth Rate, Given Birth Two

Years Since Second Birth

LEGEND

.... 1950 Cohort

– – – 1940 Cohort
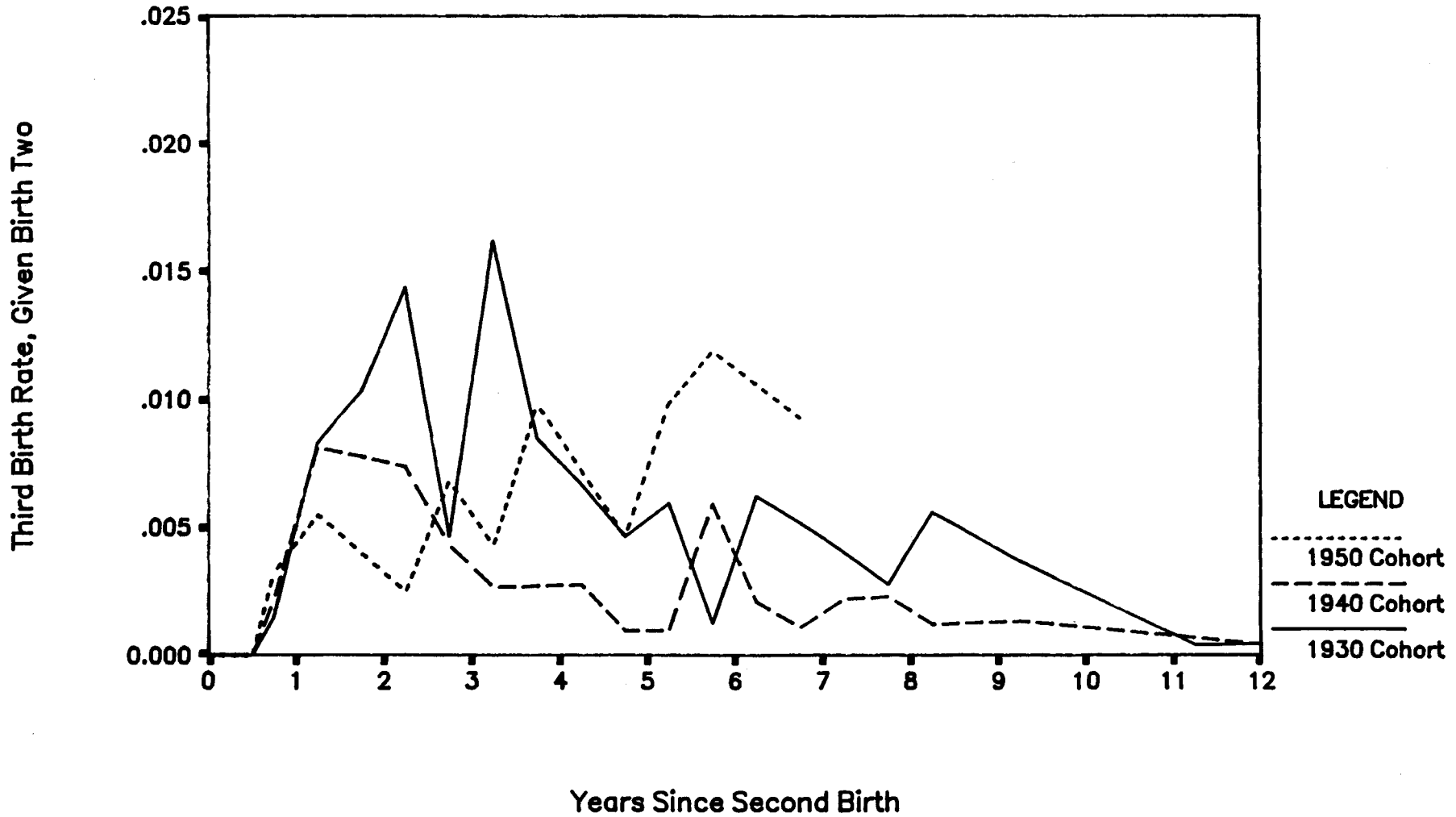
——— 1930 Cohort

1981–82 German Life History Survey

Figure 10 THIRD BIRTH RATE VS DURATION SINCE SECOND BIRTH
WOMEN BY COHORT
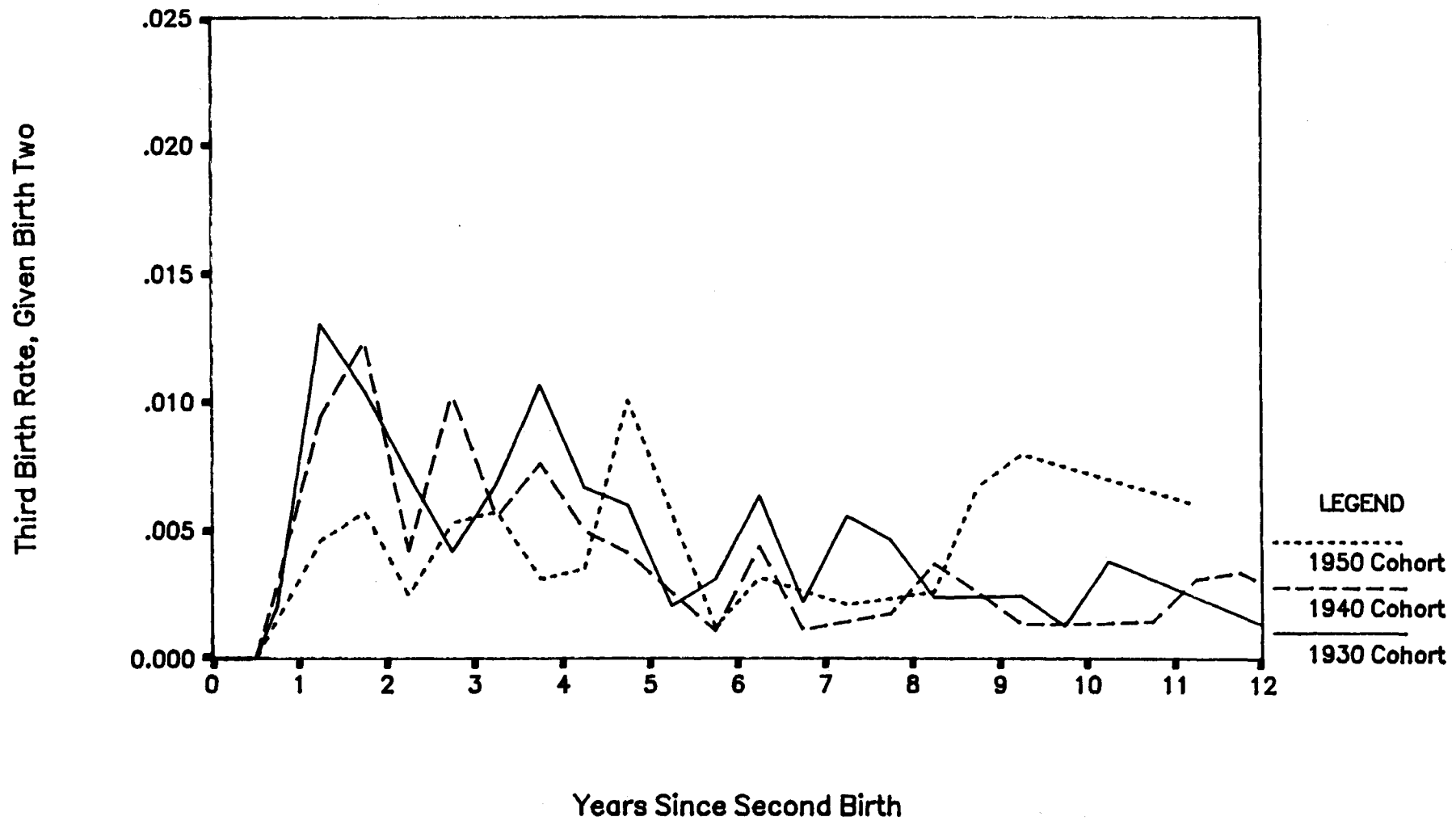
Third Birth Rate, Given Birth Two (y-axis)

Years Since Second Birth (x-axis)

LEGEND

.... 1950 Cohort

– – – – 1940 Cohort

——— 1930 Cohort

1981—82 German Life History Survey

Survivor Plot
First Birth: Men and Women, 1930 Cohort

1981-82 German Life History Study
K01S0C1
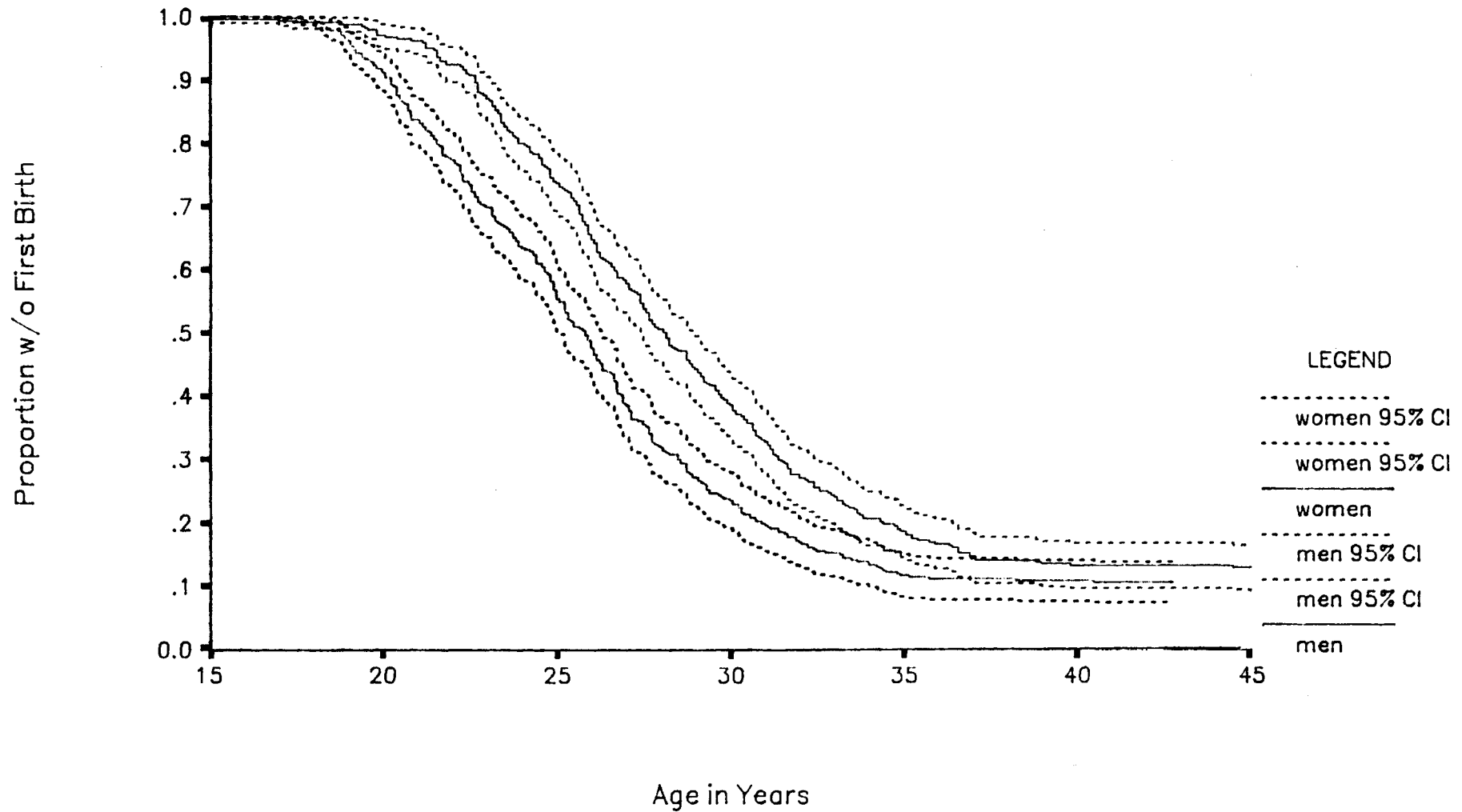
Figure A-2

Survivor Plot
First Birth: Men and Women, 1940 Cohort

Proportion w/o First Birth

Age in Years

LEGEND

women 95% CI

women 95% CI

women

men 95% CI

men 95% CI

men
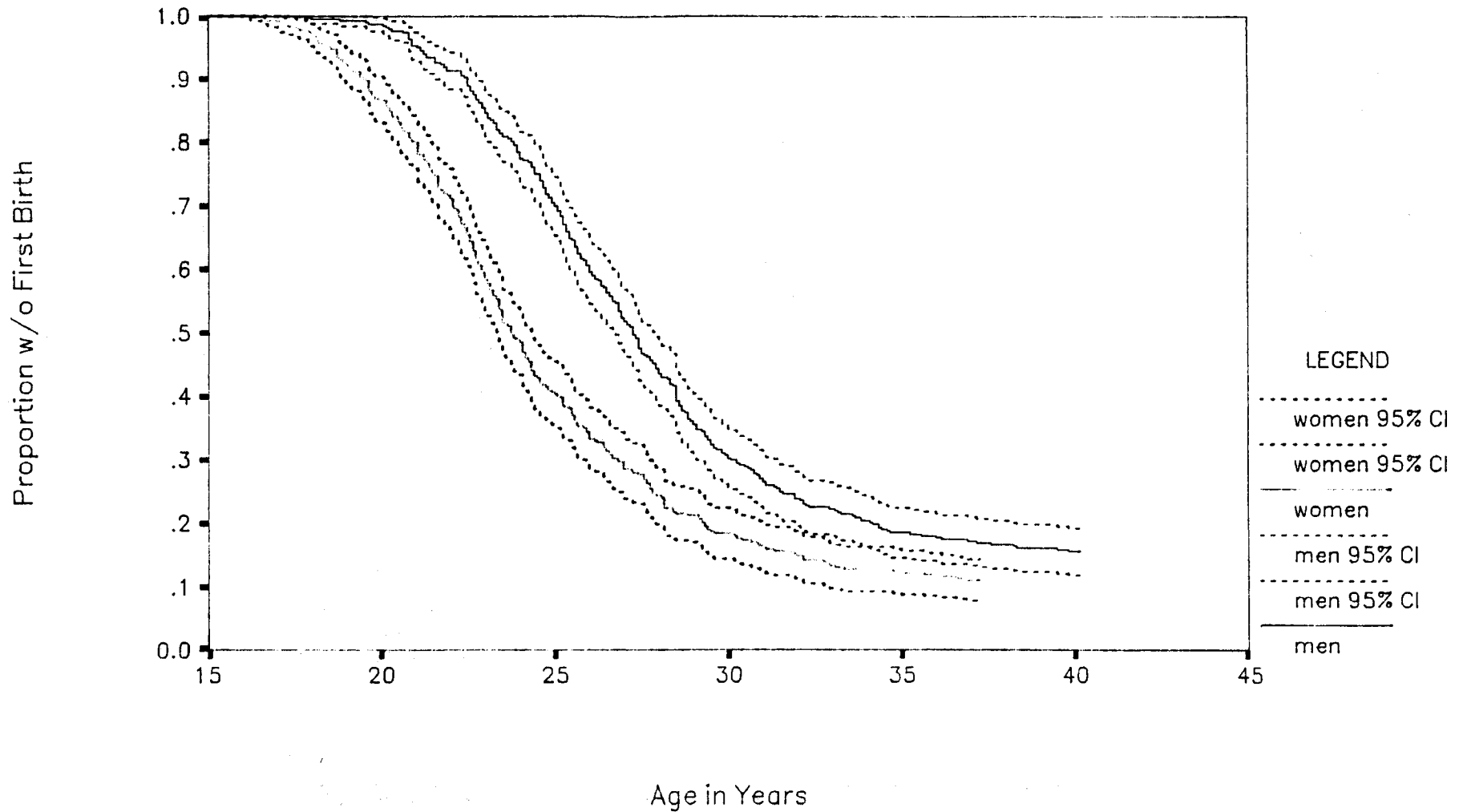
1981-82 German Life History Study
K01S0C2

Figure A-3

Survivor Plot
First Birth: Men and Women, 1950 Cohort
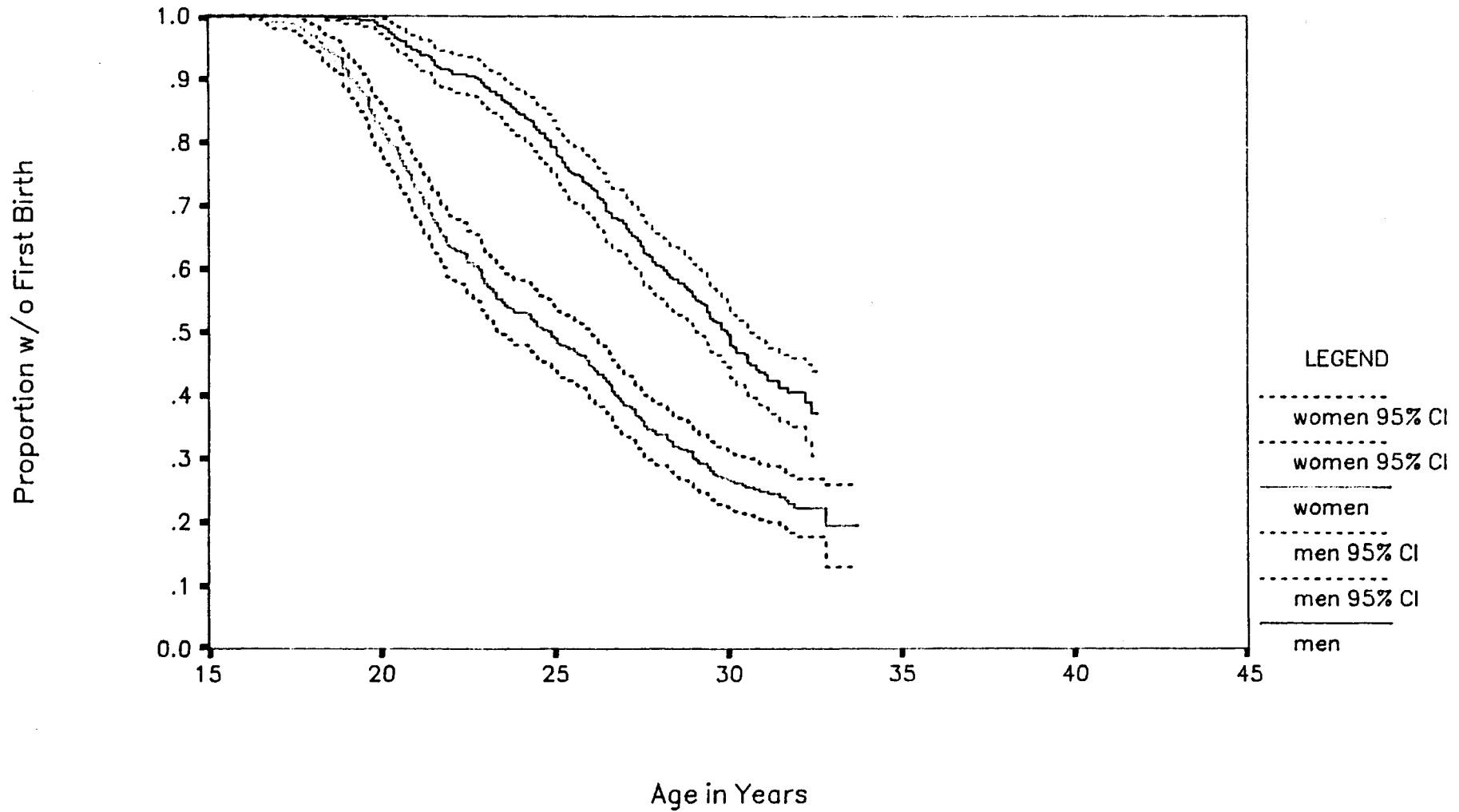
Proportion w/o First Birth

Age in Years

LEGEND

.... women 95% CI

-------- women 95% CI

———— women

.... men 95% CI

-------- men 95% CI

———— men

1981—82 German Life History Study
K01S0C3

Survivor Plot
Second Birth: Men and Women, 1930 Cohort

LEGEND

- - - - - - - women 95% CI

———— women 95% CI

———— women

·········· men 95% CI

············ men 95% CI

———— men

Proportion w/o Second Birth

Age in Years

1981—82 German Life History Study
k12s0c1
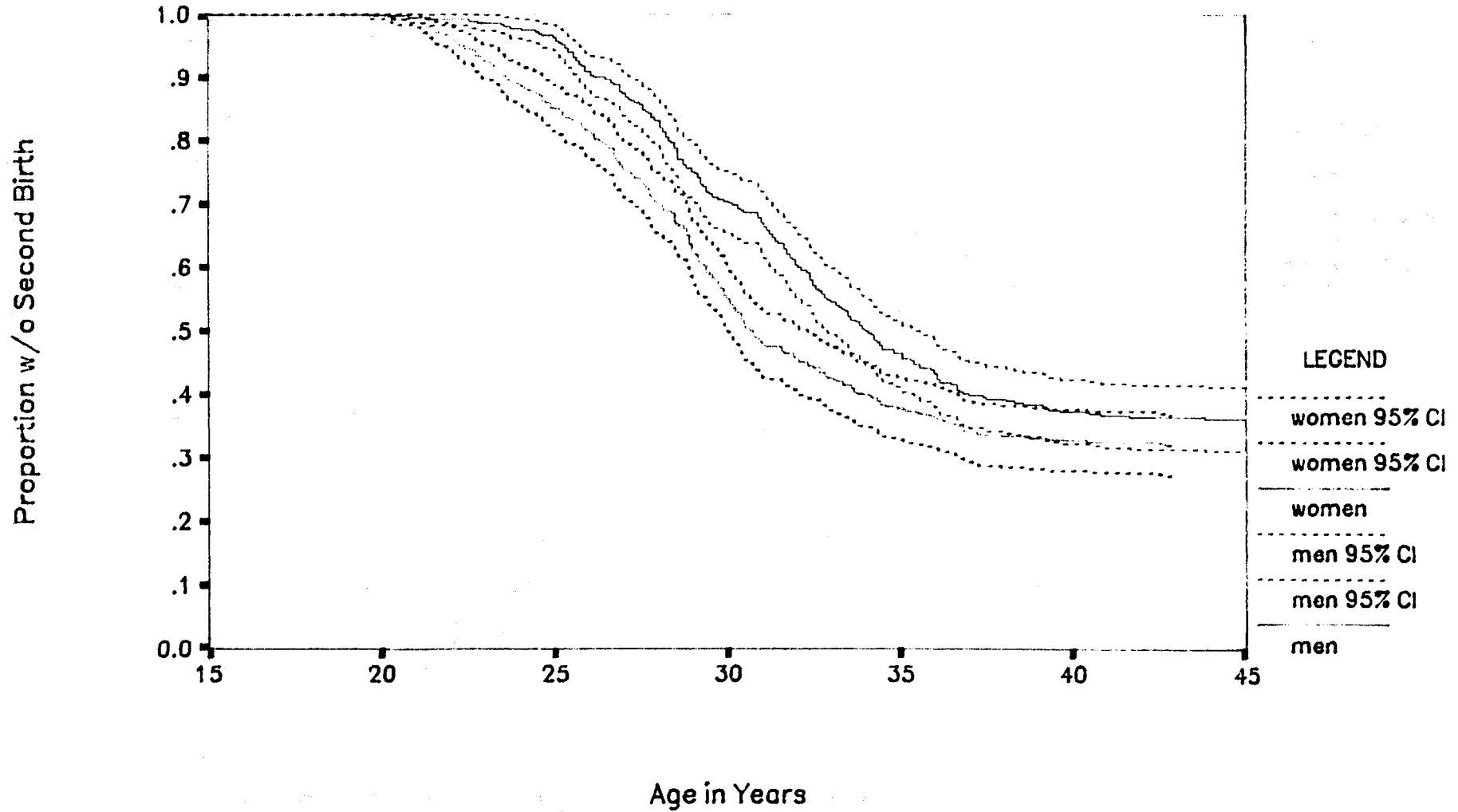
Figure A-5

Survivor Plot
Second Birth: Men and Women, 1940 Cohort
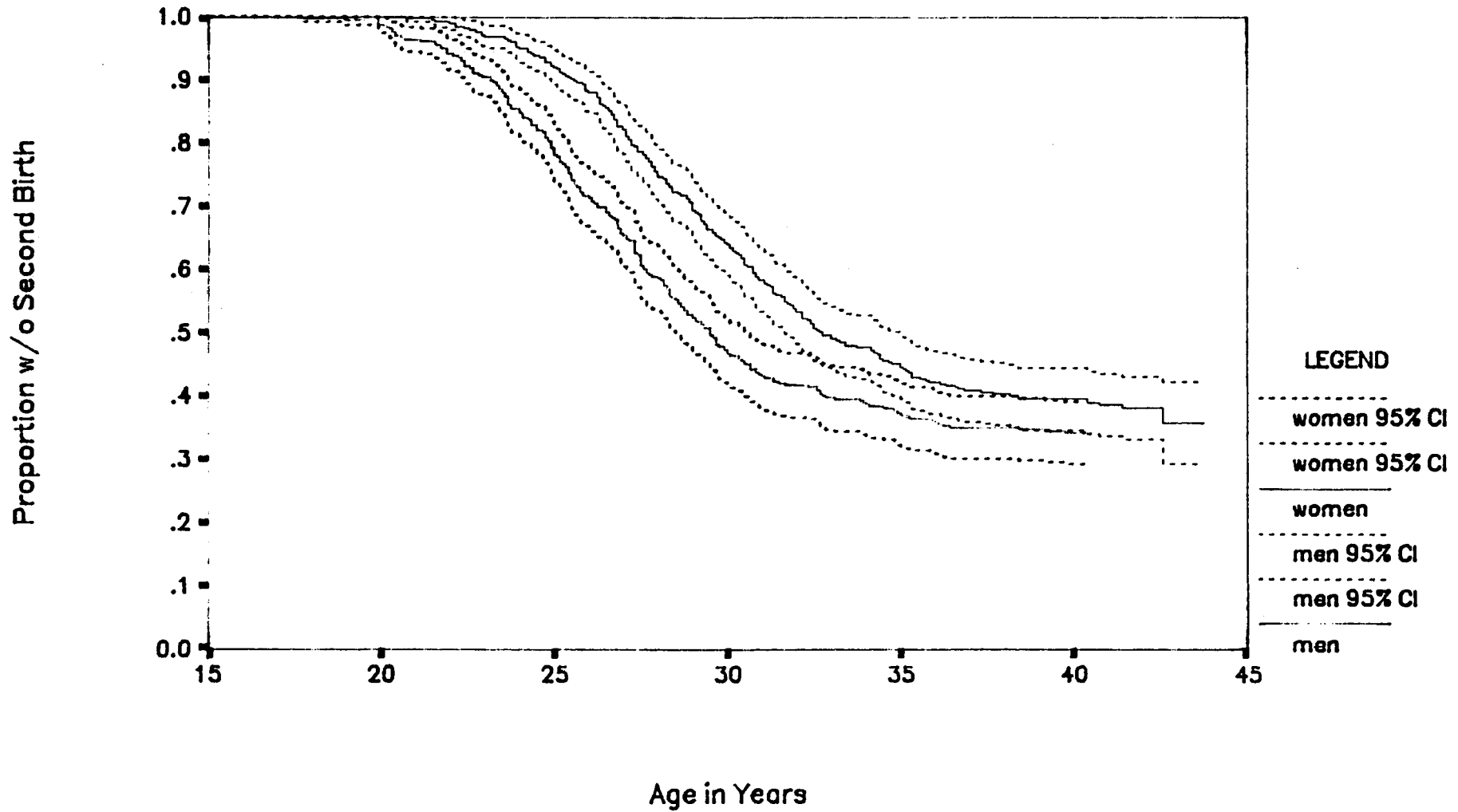
Figure A-6

Survivor Plot
Second Birth: Men and Women, 1950 Cohort
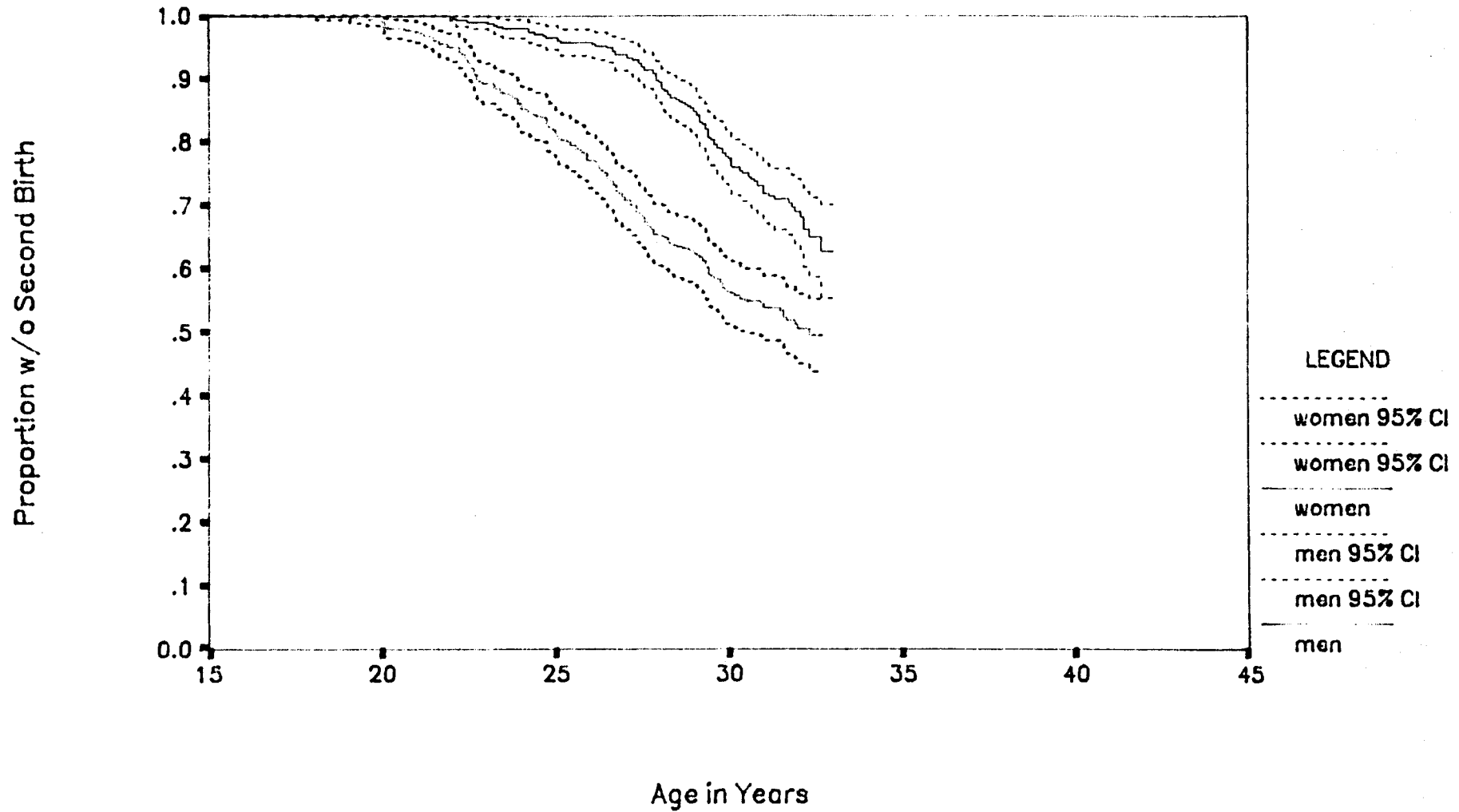
LEGEND

women 95% CI

women 95% CI

women

men 95% CI

men 95% CI

men

Proportion w/o Second Birth

Age in Years

1981—82 German Life History Study
k12s0c3

III Methodological issues

Restriction biases in the analysis of births and marriages to cohabiting women from data on the most recent conjugal union only

Jan M. Hoem, Bo Rennermalm, and Randi Selmer

1. INTRODUCTION

1.1. General problem. In retrospective studies of individual life histories, some or all of the information pertaining to periods prior to the most recent event interval will sometimes be left out. This simplifies data collection and it may improve the reliability of the data actually obtained by reducing misreporting. Unfortunately, failure to collect information on selected life history events may give rise to distortions in the analysis of marriage, childbearing, employment, and other population processes on the individual level, as is evident from contributions scattered throughout the literature.

An example of immediate concern to demographers are the effects of restricting anlyses of birth interval components to data on the latest open and latest closed birth interval only. Hobcraft and Rodriguez (1980) and Page et al. (1980) have demonstrated that such analyses can yield highly misleading results.[1] Much more needs to be known about restriction biases, however, and surprises still seem to be in store. For instance, Rindfuss, Bumpass, and Palmore (1985) recently found that latest-birth-interval restrictions did **not** seriously distort parameter estimates in logistic regressions which they made to examine determinants of birth intervals in Korea.

This paper presents the outcome of an empirical investigation into the extent of selection biases in a retrospective study of non-marital cohabitation, marriage and childbearing when information on consensual unions is essentially restricted to the most recent cohabitation and is obtained only for women who are separated, divorced, or married, or who are unmarried but

cohabit at interview. Information of this nature has apparently been collected in several surveys. The British General Household Survey is one of them (see, e.g., Brown and Kiernan, 1981), and the Danish fertility survey of 1975 is another. Our investigation was induced by problems encountered in the analysis of the latter, which can serve as a good example of its category. The general tenor of the findings presented below is that there certainly are appreciable biases in estimates based on the Danish observational plan, but that the biases need not be serious enough to preclude valid conclusions about general trends and general levels in analyses of the life-table type.

1.2. Our investigation. In the Danish survey, women who were currently married, separated or divorced when interviewed (4 071 cases) were asked about the latest marriage alone, as well as about any cohabitation with the husband just before that marriage, including the starting date of any consensual union and of the marriage. The 119 women who reported that they were separated or divorced and currently living with a man other than their latest husband, were asked to report the starting date also of their current union. Among the 1 103 unmarried respondents, the 428 women who said they were currently cohabiting, were asked to state when they started living together with their man. The 66 widowed respondents were not asked any questions about dates of cohabitation or marriage. Information about any other period(s) of cohabitation before the latest marriage or then current consensual union was not collected for any woman. Complete childbearing histories were collected for all respondents.

Various analyses of these data have appeared already; see Hoem and Selmer (1984) and their references. This paper concentrates on marriages and first births to nulliparous unmarried cohabiting women, and presents results from the Danish data in Section 2 below. Our findings indicate a dramatic fall in cohabitational nuptiality over the cohorts involved.

By contrast, the level of cohabitational first birth rates computed is remarkably stable cross cohorts. (Given the impression of skyrocketing

numbers of nonmarital births, one might have expected these rates to increase over cohorts.) A comparison with corresponding rates for married women, reported by Hoem and Selmer (1984), show that first birth rates are much lower in consensual unions than in marriage.

Because the cohabitational histories are so incomplete, these results cannot be taken at face value without an assessment of the selectivity effects inherent in the Danish observational plan. Fortunately, subsequent fertility surveys in Norway in 1977 and in Sweden in 1981 offer opportunities for checking on our findings for Danish women. Both the latter collected complete cohabitational and marital histories for all respondents along with their childbearing histories. If we disregard national differences in matters of detail and discount some national differences in process levels, they demonstrate the same general cohort trends in marriage rates and first birth rates among cohabiting nulliparous unmarried women in Norway and Sweden as we have found for Denmark. By analogy, this lends credence to our findings for Danish women despite the less fortunate observational plan of the Danish survey.

The trends mentioned for Norway have been reported by Selmer (1984); Hoem and Rennermalm (1985) have presented results concerning the behavior of cohabiting unmarried women in Sweden; and a comparison with corresponding married women can be based on reports by Finnäs (1982) and by Qvist and Rennermalm (1985). Several glimpses of these Scandinavian comparisons are given below, but a detailed documentation is not needed here. Our assesment of the Danish biases does not depend on these analogies, nice as they are, but on the following empirical experiment, recounted in Section 3.

Given access to the anonymized individual data from the Swedish fertility survey of 1981, we have imitated the ascertainment method of the Danish survey by removing from the Swedish data information collected according to the Swedish but not the Danish design, and have compared the

outcomes of the corresponding computations with the counterpart outcomes based on the full Swedish data. This has provided estimates of the biases which would have been produced if the Danish observational plan had been used in Sweden. Our general conclusion from this experiment is that for the analysis of marriages and first births to women in consensual unions, the Danish observational plan produces appreciable biases. However, these biases do not cover up or mask any trends across cohorts or any differences by age at the start of the consensual union in our Swedish data.

Like most such empirical experiments, our results are strictly tenable for our complete data set only, and its extension to other data sets, including to the restricted Danish one which induced our experiment in the first place, must be based on analogy and on informed judgement. Factors contributing to the biases are discussed at some length in Section 2 below, and we have given them further consideration in a previous mathematical paper (Hoem, 1983). Real rates of marriage and of union dissolution for cohabiting women are important contributors to the biases, and any differences in these and other factors will contribute to differences in biases in the Swedish and the Danish population. There will be similar differential effects of any differentials in nonresponse. Evidently, behavior is not identical in the two populations, and the biases in the Danish data are not exactly those estimated in our experiment. Nevertheless, in our judgement the biases operating on our Danish estimates are hardly likely to be so radically different in size or direction from those revealed by our analysis that they seriously affect a comparison between the two countries. We find it even harder to believe that differential biases can give any important distortion of the trends found below in marital and childbearing behavior over our cohorts. Despite their differences, demographic processes in the Danish and Swedish populations must be among the most closely similar anywhere.[2]

In this spirit, a comparison of the outcomes of the two surveys offers the insight that nulliparous Danish women living in a consensual union marry more readily and are much less prone to get a child outside marriage than are their Swedish counterparts in each comparable cohort. By this token, Sweden is a trend setter, for together the two populations are world leaders among industrialized nations in the prevalence of modern nonmarital cohabitation.

To summarize this introduction, this paper has several foci. Its main contribution is the assessment of restriction biases in Section 3. The outcome of this assessment is that the empirical investigation of the Danish data reported in Section 2 gives somewhat biased but largely valid results. Because it does, a brief comparison of the relevant demographic behavior in Denmark and in Sweden can be given in Section 4.

2. COHABITATIONAL MARRIAGES AND FIRST BIRTHS IN DENMARK

2.1. The Danish data.

The Danish fertility survey of 1975 achieved interviews with 5 240 respondents, which corresponded to a response rate of 88 per cent. Selective nonresponse is unlikely to have had an important effect on our results. As we have noted, however, there may be selection biases due to the fact that information was not collected on periods of cohabitation before any latest marriage or current consensual union. This lack of knowledge of a woman's cohabitational status at any time outside of the periods actually recorded, precludes any analysis of noncohabitational births and rules out almost all analysis of the formation of consensual unions and of marriages not starting in such unions. This paper shows that it is possible to make a sensible investigation of some of the behavior in consensual unions by means of life table methods.

The questionnaire left it to the respondent to define whether she was or had been living in a consensual union as described. This seems to be common practice, and presumably the situation perceived by the respondent herself is as good an operational definition as any for our purposes. Nevertheless, there must have been some problems of reporting accuracy of a status of such uncertain definition, and which some women may have been reluctant to reveal, although few real problems seem to have arisen in the interview situation. In particular, many women will have found it difficult to pinpoint a starting date or to remember it

accurately when interviewed, even though only the month and year was asked for this date as well as for all other dates of events in the interview. This shows up in the relatively high nonresponse rate for such starting dates (246 cases out of 1 571 recorded nonmarital unions), and also in an inflated number of reported starting dates about a half and a whole year before marriage. Such digital preference is revealed in our study of cohabitational marriage rates, reported in Section 2.3 below, but it cannot have distorted our main results in that section or elsewhere, for the results do not depend on details of this nature.

For a further discussion of the reliability of Scandinavian interview data on cohabitation, see Hoem and Rennermalm (1985). Further information about our data was given by Finnäs and Hoem (1980) and by Hoem and Selmer (1984).

As a move towards homogeneity of each population group and with a view to removing disturbing effects of differential behavior of unmarried women who start the recorded cohabitation with previous children or previous marriages, we have excluded from analysis all consensual unions with a starting parity above 0 and a reported starting age of 25 or more. Remaining consensual unions have been classified according to the reported age of the woman at the start of the union, into those who started at an age below 20 and those who started at ages 20-24. This is to control for differential behavior by age at initiation. Similar birth cohort differentials are picked up by grouping our respondents into conventional five-year birth cohorts born in 1926-1930, 1931-1935,..., 1951-1955. Women who reported premarital cohabitation but gave insufficient in-

formation about their starting age in this union, were deleted, and so were a negligible number of further records because some other vital information was missing or evidently wrong, as well as some respondents born in 1925, 1956, and 1957.

After these various deletions, we have ended up with 1 240 women whose records satisfy our present requirements, which are that they should contain information, not irreparably deficient, on a consensual union started at parity 0, before any first recorded marriage, and at an age below 25, for a woman born in 1926-1955. We should perhaps emphasize that our present analysis deliberately is restricted to the data on these 1 240 women. We do not draw conclusions about population totals or population means, valid, say, for all of the Danish population at the time of the survey or at any previous time, as in a standard enumerative survey. Instead, we take the superpopulationist stand relevant to life table analysis. A further discussion of these matters was given by Hoem and Selmer (1984). The deleted records are not relevant for this analysis and are not documented here. Sizes of the groups used here can be seen in our Table 1.

In a few cases where the starting year of a nonmarital cohabitation was reported but the starting month was missing, and where the woman had a cohabitational first birth no earlier than in the subsequent year, a starting month of July was imputed to permit the approximate calculation of starting age. These cases were not included in the analysis of the first birth and of marriage formation for nulliparous unmarried cohabiting women, but they were taken into account in our analysis of marriage formation after the first birth (Table 1), reported in Section 2.4.

Table 1. Number of women who reported premarital cohabitation and stated its starting date in the Danish fertility survey of 1975. By cohort and by starting age. Women with starting parity 0 only.

Cohort born	Number	Of these: Married before first birth	Nonmarital first births	Married after first birth	Nonmarital second births
	(1)	(2)	(3)[a]	(4)[b][c]	(5)[b][d]

Age below 20 at reported start of consensual union

1926-30	17	16	1 (+1)	2	0
1931-35	28	23	5 (-1)	4	0
1936-40	58	54	4 (+2)	6	0
1941-45	100	82	16 (+7)	18	5
1946-50	147	117	27	24	0
1951-55	223	91	33 (+1)	15	2
	573				

Age 20-24 at reported start of consensual union

1926-30	20	18	2	2	0
1931-35	28	24	4 (+2)	3	3
1936-40	59	54	3	3	0
1941-45	139	120	15	13	1
1946-50	261	202	20 (+4)	10	2
1951-55	160	40	12 (+1)	3	0
	667				

a. Positive figures in parentheses after Column 3 are numbers of women who reported only the starting year of unmarried cohabitation, and who also reported a nonmarital first birth no earlier than the subsequent year. The negative figure -1 in parenthesis for the cohort born 1931-35 is for a woman who had twins (among the five giving nonmarital first birth). The sum of a figure in Column 3 and the corresponding figure in parenthesis is the number of women counted as exposed to subsequent risks after a first single birth in the relevant cohort.

b. Numbers out of those exposed to this risk, mentioned at the end of footnote a.

c. Numbers married before the interview, between a first birth and any second birth.

d. Out of the 13 cases of this column, only three were unmarried at interview.

2.2. Methods of analysis. Our investigation uses straightforward life table methods, in this case with two decrements (marriage and first birth). Life table methods are particularly well suited for the analysis of waiting time data with heavy censoring, such as ours. Each group listed in Column 1 of Table 1 has been followed month by month from the beginning of cohabitation, recorded as starting in ordinal month 0 of the union since dates were only recorded by month and year and ordinal months were computed as differences between calendar months. Respondents have been recorded as decrements from observation at marriage, at first births, or at interview (censoring), timed to happen in the middle of the ordinal month of occurrence. For each ordinal month, an ordinary occurrence/exposure rate has been computed for recorded marriages, as well as another one for first births.[3] As is evident from Table 1, some groups simply are too small for analysis by this method, and even for the larger groups, attrition is often so rapid that it is hard to study behavior over longer durations than a couple of years. See Figure 1 for a case in point.

2.3. Results on cohabitational nuptiality in Denmark. Scandinavia has an old-standing tradition of nonmarital cohabitation (Hyrenius, 1941; Trost, 1978; Matović, 1985), and our data contain a lot of evidence of the recent strong extension of its prevalence. Our computations of marriage rates for nulliparous women by single months of duration of reported consensual unions add colour to this picture. Marriage rates computed for Danish women who reported an age below 20 years at the start of cohabitation
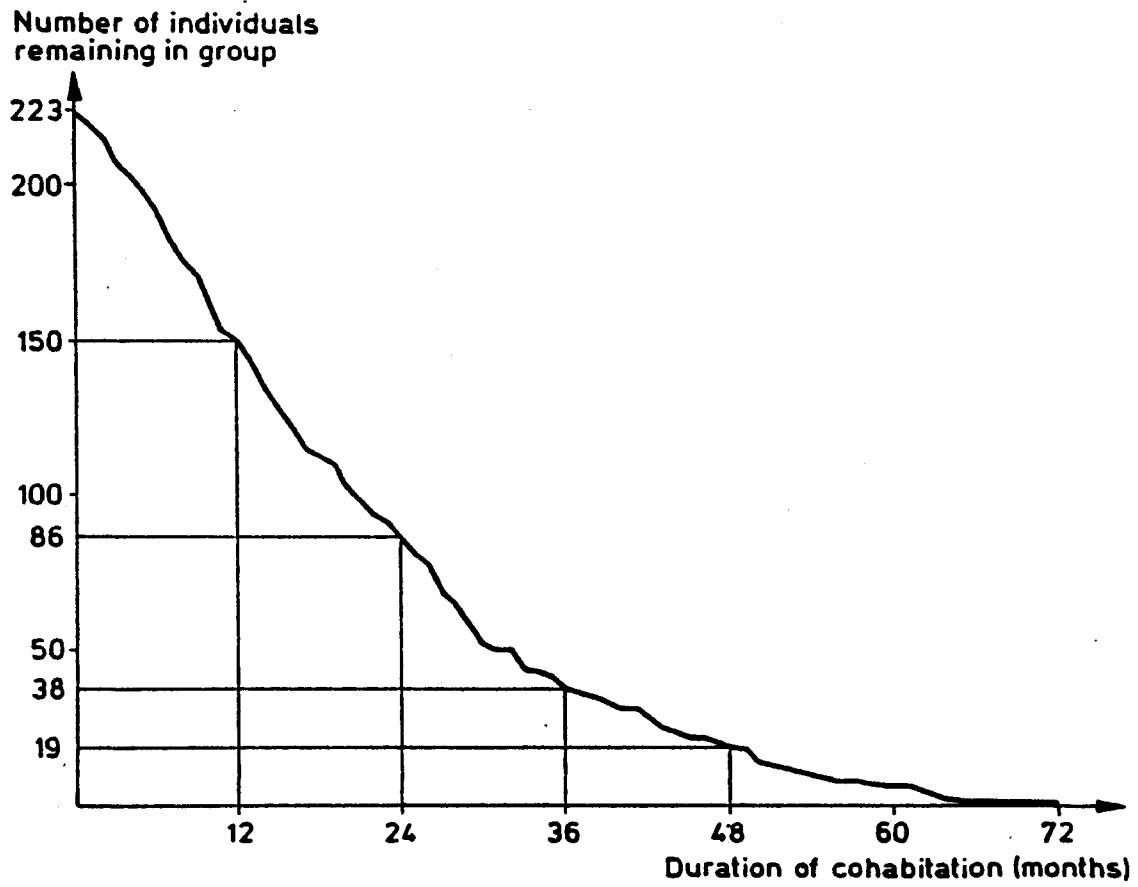
Figure 1. Group attrition by duration of nonmarital cohabitation.
Danish cohort born 1951-55, starting cohabitation at
ages below 20.

are a bit lower in each cohort than corresponding rates for those who said their union began at age 20-24 (Table 2, Columns 1 to 3), but the difference has seemed relatively unimportant, so we have combined the two groups in each cohort (Table 2, Columns 4 to 6).

A plot of the cohabitational marriage rates then typically is as in Figure 2, which is for the cohort born 1946-1950. We regard some features of the diagram as reflections of genuine properties of the underlying intensity function, others as artifacts due to reporting errors or the influence of sampling variations, as follows.

The increase in the marriage rates over the first few months of cohabitation probably reveals a corresponding increase in the intensity function. A similar feature appears generally in our plots of marriage rates for the various groups studied, and it is easily explained in terms of some initial waiting time needed before partners in a newly established union can get around to becoming married.

On the other hand, the peaks at durations of 6 and 12 months in Figure 2 are indications of digital preference in the reporting of starting months of cohabitational unions leading to marriage. Evidently, many women in this category have "rounded off" the reported duration of the premarital union to half a year or to a full year. More accurate reporting no doubt would have placed many of these unions at neighbouring durations. Because of this, and since we can see no obvious trend in the cohabitational marriage rates over durations above the first few months, we have recombined the data and have computed a common rate for the interval from 5 to 23 months, with results given in Table 2. (We could have gone above a duration of 23 months in the data of Figure 2, but not in several other, smaller groups involved. Note the uneven size of the exposures in Table 2.)

Table 2. Marriage rates for nulliparous Danish women in consensual unions,
at durations of 5 to 23 ordinal months combined. By cohort and by
age at start of union. Per 1000 women per month.

Cohort born in	Marriage rate for starting ages		Proportionate increase from (1) to (2)	All starting ages below 25			
	below 20	20-24		Occur-rences	Exposures (months)	Marriage rate	Partial mar-riage probability (per cent)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1926-30 ⎫				16 ⎫	178.5 ⎫		
1931-35 ⎬	71.4	87.8	23	20 ⎬ 89	288 ⎬ 1126	79	78
1936-40 ⎭				53 ⎭	659.5 ⎭		
1941-45	66.6	73.1	10	111	1582.5	70	74
1946-50	49.3	50.0	1	182	3657	50	61
1951-55	19.7	22.0	12	74	.3629	20	32
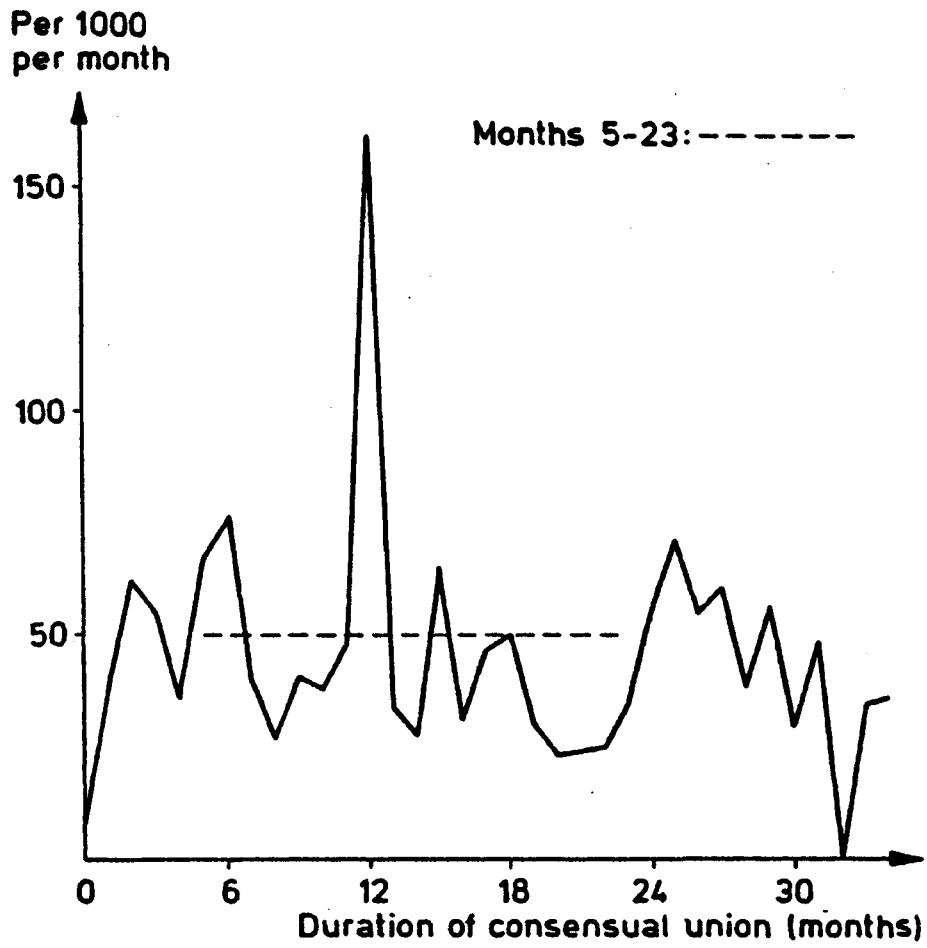
Figure 2. Marriage rates for nullipara by reported duration of con-
sensual union starting at all ages below 25. Danish
cohort born 1946-50.

One may possibly get a better feeling for the import of numerical values of rates like those in Column 6 of Table 2 by transforming them into the domain of probabilities. Column 7 contains the value of $1 - \exp\{-19\hat{\eta}\}$ for each marriage rate $\hat{\eta}$. The transformed value is an estimate of the probability of marrying before the union has lasted for two years, for a nulliparous woman who has lived in a consensual union for five months, in the absence of any competing risks (death, union dissolution, first birth). According to Table 2, this probability has been about halved between our two youngest cohorts.

The tremendous drop in the cohabitational marriage rate observed, particularly in the two youngest cohorts, needs to be interpreted with some care, for it may be influenced both by reporting errors and by biases due to the observational plan. A glance at Table 1 shows that in the cohorts born in 1950 or before, most consensual unions reported lead to marriage, while in the youngest cohort, a large proportion of the cohabiting women had not had time to marry yet, or were in a reported union which might dissolve at some date after the interview. Perhaps women in the other cohorts remembered or revealed fewer of the latter kind of unions. Some of the recorded drop may be due to reporting errors of this nature, which would tend to overstate a real decrease.

Even if reporting were perfect, the fact that our analysis is based on information obtained about the latest consensual union only (before marriage or interview) may have led to biases on two counts. First, in reality, consensual unions differ in nature, and whether one eventually turns into a marriage will depend on its characteristics, such as its stability. Our observational plan may well have led us to selectively overrecord unions of a nature likely to lead to marriage in the older cohorts, but not so much in the younger ones. This would reinforce a tendency to overdramatise a real decrease.

Secondly, the practice of recording the latest union only will have led to consistent overestimation of real cohabitational nuptiality in each cohort. Assuming for this argument that all marriages in our analysis are first marriages, and also assuming that all unions in a cohort have roughly the same nuptiality, the occurrences of the risk considered will have been recorded correctly but time spent in any consensual unions before the ones reported will be missing from the exposures, leading to inflated nuptiality rates. The extent of this upward bias will depend on several factors. For instance, it depends strongly on the level of risk of dissolution of a consensual union. If this risk has increased (say) from older to younger cohorts, corresponding to increasing casualness in cohabitational relations and parallel to increasing divorce risks in marriage, then the overestimation will be progressively greater in younger cohorts, ceteris paribus. This would tend to mask a real drop, which would be stronger than the one recorded if the effect of all other factors remained the same across cohorts.

This effect would be diluted by the fact that young women just would not have had the same time to experience periods of cohabitation prior to the one recorded at interview as older women would have if the "risks" of forming an dissolving cohabitational unions were the same at each age. If other things were equal, therefore, less exposure time would be missing for the younger women.

Some other aspects of the observational design may have influenced occurrences as well as exposures. Widows should have contributed to occurrences and exposures to the extent that they lived in consensual unions before first marriage, but they now contribute nothing to either. Furthermore, we cannot be certain that all recorded marriages are first marriages, and a remarried woman will contribute wrong amounts to the exposures and possibly also to the occurrences if after some marriage

she enters a recorded consensual union with no prior children and
then marries her cohabitant. Her recorded exposure will then come from
her recorded cohabitation rather than from whatever time she spent in
consensual unions before her first marriage. If she has her first child
in the recorded consensual union, she will also contribute to recorded
first births, when if anything she should have contributed to first mar-
riages. A woman who has her first child in one consensual union and gets
a subsequent union recorded by us, similarly will wrongly contribute
nothing to the exposure and also will not contribute to the first-birth
occurrences, as she should.

Many of these items may seem trivial, and one might at first assume
that the influence of several of them must be negligible. Some of our
groups are so small, however, that incorrect recording of data about
single women caused by the observational plan turn out to be most in-
fluential. The total outcome of the various conflicting effects cannot
be assessed accurately from general reasoning alone; external information
is needed, and is presented in Section 3. One cannot get sufficient in-
sight from the flawed data themselves. It will not do, for instance,
to compute a marriage rate separately for those who actually marry in
each cohort, for such a rate will be strongly influenced by other
selection biases discussed elsewhere in the literature.

This discussion of potential biases in comparisons between cohorts
reveals the possibility that there may be similar biases in the com-
parison made between the cohabitational nuptiality rates by starting
age for each cohort. Just as the older cohorts have had more time to
lose exposure in previous consensual unions, women reporting a starting
age of 20 to 24 years for their current union may have more unrevealed
exposure than those in the same cohort who reported a starting age below
20 years. This may have led the cohabitational nuptiality rates

of the older starters to be more overestimated than for younger starters.
Our computed rates are only weakly dependent on reported starting age,
but according to this argument, some tendency of older starters to
marry more slowly than younger starters may be covered up by differ-
ential biases. On the other hand, some of the women for whom we have
recorded a later starting age, may actually have an unrevealed earlier
real debut into cohabitation. Since the corresponding unrevealed ex-
posure is missing in the marriage rates computed for starting ages
15 to 19 years, this would work towards distorting the latter rates
upwards. What the balance is between these competing biases can only
be assessed empirically, as in Section 3 below.

2.4. Results on first births to unmarried cohabiting Danish women,
and on subsequent marriage. As can be seen in Table 1, rather few women
have reported first births in consensual unions. The risk of such first
births must have been quite small, as is evident from Table 3 as well.
We are unable to detect any real trend in our rates from older to
younger cohorts. Formal significance tests (see Hoem and Selmer, 1984)
suggest that the cohort differences seen in Table 3 may well be due
largely to random fluctuations. On the other hand, the tendency for
the first birth rate of the members of a cohort who reported a start-
ing age below 20 years, to be somewhat higher than the corresponding
rate for those who said they started at ages 20 to 24, seems to re-
present a significant difference. As in many other contexts, those who
report an earlier start, appear to have a higher pace of fertility at
this stage as well.

The difference discovered here may possibly even underrate a real
difference. Since early periods of cohabitation may have been missed

during data collection so that we cannot be certain whether the starting age recorded is the real age at debut of cohabitation, some of those assigned to the later starting age group may actually belong in the earlier group for age at real start of cohabitation. If the cohabitational fertility of such women is on a par with those whose real cohabitational debut is at ages 20 to 24, then they are classified correctly by our system, and the birth rates computed would reveal real differences (apart from random fluctuations and in the absence of other biases). If fertility depends on real rather than recorded starting age, they would be misclassified by our system, and a correct classification would tend to reveal a greater real difference than the one observed by us. Intermediary levels of fertility for this group would also tend to underrate the difference.

Table 3. First birth rates for Danish women in consensual unions of a duration up to 24 months. By cohort and by reported age at start of cohabitation.

Cohort born	Ages below 20 at start of cohabitation		Ages 20-24 at start of cohabitation	
	Exposures (months)	Rate per 1000	Exposures (months)	Rate per 1000
1926-30	164	6.1	160.5	12.5
1931-35	242	16.5	265	15.1
1936-40	624.5	6.4	538.5	5.6
1941-45	1152	11.3	1454.5	6.9
1946-50	2039	10.8	3473	2.6
1951-55	3640.5	6.9	1759	6.8
All cohorts	7862	8.8	7650.5	5.2

Otherwise, the first birth rates computed here are subject to much the same kind of biases as those discussed for the marriage rates of women in consensual unions. As noted, there could be some contamination of our sets of presumed premarital unions by unrevealed intermarital unions for divorced women who later married and only got to report the second marriage. Cohabitational first birth rates have the same problem with the exposures as cohabitational marriage rates have, for the same computed exposures are used for both kinds of rates. In addition, cohabitational first births may be undercounted in the Danish system, because some such births occur in consensual unions before the ones analysed by us. This will tend to counteract the upward bias due to the underestimation of exposures. A first birth rate for consensual unions computed here will tend to be less of an overestimate of the real rate than is the case for the corresponding marriage rate. In fact, so many first births may be missing that a computed first birth rate may be an underestimate of the real thing. Again, an empirical investigation such as that of Section 3 is needed to see how these conflicting effects work themselves out in practice.

Because of the interaction of competing risks, constant first birth fertility and falling nuptiality in consensual unions imply an increasing probability of having a first birth in the union during a given period under constant mortality and constant risk of union dissolution. To illustrate this effect by a "pure" measure, let us disregard the latter two risks in the usual manner, and let us make calculations as if consensual unions can only convert into marriages. For a constant first birth intensity α and a constant marriage intensity η, the probability of having a first birth during a period of length z is then

$$\Pi(z) = \int_0^z e^{-(\alpha+\eta)t}\alpha \, dt = \frac{\alpha}{\alpha+\eta} (1 - e^{-(\alpha+\eta)z}).$$

If we take the marriage rates in Column 6 of Table 2 and the birth
rates for all cohorts combined in Table 3 at face value, then esti-
mates of $\pi(z)$ for each cohort and each group of starting ages can be
computed as in Table 4. We have used z=19 months because Table 2 was
based on data for a total of 19 months of union duration. By Table 4,
for each starting age group the estimated probability has increased
by some 60% between the oldest cohorts and the youngest one. The final
line in Table 4 shows what the estimate $\hat{\Pi}(19)$ would be if there were
no cohabitational nuptiality, i.e., if $\eta=0$.

Table 4. Estimated probability $\hat{\Pi}(z)$ of having a nonmarital first
birth in a consensual union over a period of z=19 months
under the competing risks of childbearing and marriage,
disregarding mortality and union dissolution. Percent.

Cohort born	Ages at start of cohabitation	
	below 20	20-24
1926-40 (combined)	8.1	4.9
1941-45	8.7	5.3
1946-50	10.1	6.1
1951-55	12.9	7.9
No nuptiality	15.4	9.4

Footnote: Marriage rates from Table 2, by cohort.
          Birth rates from Table 3, for all cohorts combined.

Since so few of our respondents have reported a cohabitational first birth, analysis of subsequent marriages separately by cohort or starting age has not been possible by life table methods. A graph of estimated marriage rates by time since first birth for all cohorts and both starting age groups combined (Figure 3) shows that as expected, the marriage intensity has a marked peak over the first few months after childbearing. Cohabiting women who give birth outside of marriage do indeed have a strong tendency to marry shortly afterwards. Between some six to sixteen months after childbearing, the marriage rates are rather stable, at a level similar to the marriage rates of cohabiting women without children (Figure 2). After eighteen months, the marriage rate is very small for those who have become mothers. Even though many of the new mothers tend to marry, the urge to do so has not been overwhelming in these cohorts. As computed by the single decrement life table method, some four out of every ten of them would stay unmarried beyond eighteen months of childbearing, in which case they cannot be said to have married in connection with the first birth. The partial marriage probability for all cohorts is estimated to be 0.57. This may have been deflated somewhat by a stronger tendency to stay unmarried beyond the first birth in our youngest cohort, but the estimate only rises to 0.61 when the latter is removed from the data, so the number unmarried is still roughly four in ten.
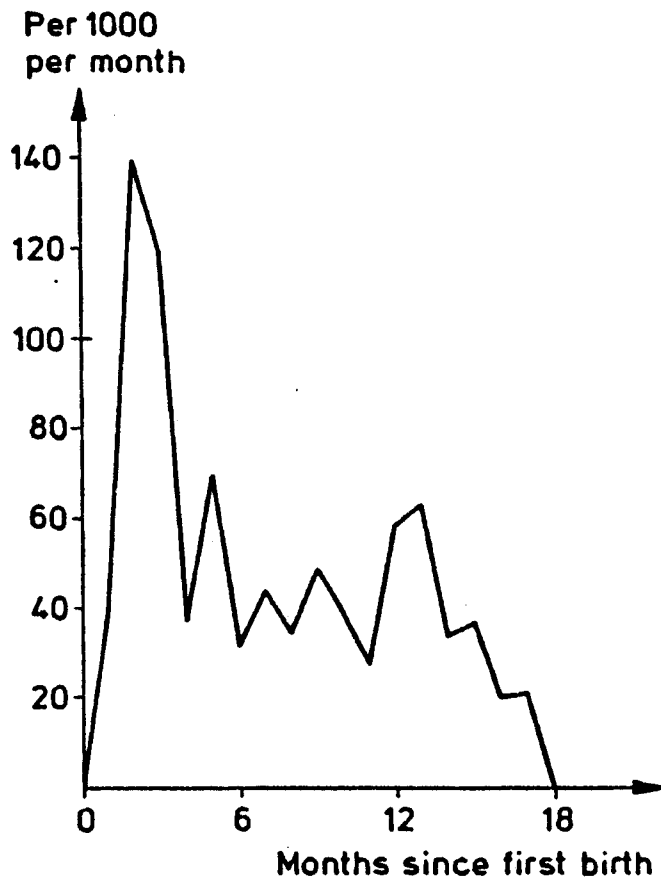
Figure 3. Marriage rates by time since first birth.
Unmarried women starting cohabitation at
ages below 25 years. All Danish cohorts combined.

3. THE EXTENT OF DISTORTION

3.1. The Swedish data. We now turn to our assessment of the biases which the Danish observational design would have produced in the Swedish data. In the aspects of interest here, the nature of the data collected in the Swedish fertility survey of 1981 are much the same as in the Danish survey of 1975, except that in Sweden complete cohabitational and marital histories were collected. The sample was drawn by simple random sampling from each of the five-year cohorts born in Sweden in 1936-1940, 1941-1945,..., 1956-1960 and registered as resident in the country at the time when the sample was drawn, irrespective of marital status. Interviews were achieved with 4 300 respondents, which corresponded to a response rate of 87%. The demographic behavior of nulliparous unmarried cohabiting women and some other aspects of these data have been analysed by Hoem and Rennermalm (1985), who also describe the data more fully. For further information about the Swedish data, see Arvidsson et al. (1982) and Lyberg (1984).

There are some differences between the two data sets, or in our treatment of them. A comparison with Section 2.1 shows that the two surveys have the four cohorts born in 1936-40, 1941-45, 1946-50, and 1951-55 in common. The Danish data also contain the two previous five-year cohorts born in 1926-30 and 1931-35, while the Swedish data have the additional youngest cohort born in 1956-60. Thus, there is a shift in coverage. It is also possible that the more complete coverage of cohabitational histories in the Swedish questionnaire may have led to greater response reliability than the corresponding more restricted formulation in the Danish questionnaire would. Given the general nature of the results of our bias assessment, presented below, it is unlikely that the shift in coverage or potentially in response accuracy can have been important for our interpretation of our experiment, however.

With our group sizes, random variation is frequently only too important
an element in our bias estimates. (Some further, trivial differences
between the data sets are discussed in Endnote 3.)

3.2. Distortion results. After selection of eligible respondents
and deletion of (a few) records with irreparably deficient information,
we have ended up with the group sizes listed in Table 5. Month-by month
exposures, occurrences, and distortion factors for the cohabitations
of one of these groups are given in Table 6 for illustration. At a
certain duration, let E_S and O_S be the exposures and one of the three
occurrences (say the number of marriages) recorded by the Swedish
observational plan, and let E_D and O_D be the corresponding exposures
and occurrences, respectively, produced by imitating the Danish ob-
servational plan on the same individual level data. Then the Danish
type occurrence/exposure rate O_D/E_D has an estimated bias of

$$\hat{b} = \frac{O_D/E_D}{O_S/E_S} - 1.$$

Bias values have been listed (in percent) in Columns 8 and 9 of
Table 6. Figures 4 and 5 contain plots of similar sequences of dis-
tortion factors $100(1+\hat{b})$ for some other cohorts. Note that single-
month duration intervals have been used for ordinal month 0 to 11,
two month intervals were applied for ordinal months 12 to 23, and
twelve-month intervals were used for subsequent ordinal months. The
gap at ordinal month 4 in the curve for the cohort born in 1956-60 in
Figure 4 is due to the fact that there were no recorded marriages
there $(O_S = O_D = 0)$, in which case the estimated bias $\hat{b}$ is undefined.

Several features of Table 6 and Figures 4 and 5 deserve comment:

Table 5. Group sizes in our experiment with data from the Swedish fertility survey of 1981.

Cohort born	Recorded number of cohabitations among (apparently) never-married nullipara					
	by the Swedish design[a], at starting ages			by the Danish design[b], at starting ages		
	below 20	20-24	all[c]	below 20	20-24	all
1936-40	59	78	189	38	69	154
1941-45	149	255	543	116	217	445
1946-50	227	417	786	157	326	594
1951-55	371	500	941	239	375	675
1956-60	346	200	546	215	165	380
Total	1152	1450	3005	765	1152	2248

a) I.e., in the complete data set.

b) I.e., by an application of the Danish observational plan to the Swedish data set.

c) Starting ages below 25 as well as 25 and over.

Table 6. Month-by-month exposures, occurrences, and bias for unmarried nulliparous women living in consensual unions, recorded (i) in the Swedish data and (ii) by applying the Danish design to the same data. Swedish cohort born 1941-45. All starting ages combined.

Duration of union (ordinal months) a)	Swedish design				Danish design			Bias (percent)		Rate vari- ability b)
	Expo- sures	Mar- riages	First births	Disso- lutions	Expo- sures	Mar- riages	First births	for mar- riage rates	for first birth rates	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0	271.5	0	0	0	222.5	0	0	-	-	-
1	526.5	27	6	0	430.5	25	4	13.2	-18.5	38.5
2	494	25	6	1	401.5	24	5	18.1	2.5	40.0
3	464	20	5	3	375.5	19	4	17.4	-1.2	44.7
4	433.5	23	7	3	350	22	6	18.5	6.2	41.7
5	401.5	25	4	2	321.5	25	4	24.9	24.9	40.0
6	370.5	22	6	3	294.5	20	5	14.4	4.8	42.6
7	346.5	11	5	1	274	11	5	26.5	26.5	60.3
8	331	10	3	1	259.5	10	3	27.6	27.6	63.2
9	318.5	7	4	0	249.5	5[c)	2[d)	-8.8	-36.2	*

Table 6. (continued)

Duration of union (ordinal months) a)	Swedish design				Danish design			Bias (percent)		Rate vari- ability b)
	Expo- sures	Mar- riages	First births	Disso- lutions	Expo- sures	Mar- riages	First births	for mar- riage rates	for first birth rates	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
10	305	10	4	1	239.5	9	3	14.6	-4.5	63.2
11	290	10	4	0	226	10	4	28.3	28.3	63.2
12-13	522	28	10	4	407.5	25	6	14.4	-23.1	37.8
14-15	459.5	17	6	·2	354.5	17	6	29.6	29.6	48.5
16-17	407	16	7	1	305.5	16	7	33.2	33.2	50.0
18-19	360.5	15	6	1	263	15	4	37.1	-8.6	51.6
20-21	322.5	11	2	2	229.5	11	2	40.5	40.5	60.3
22-23	293.5	12	2	3	206.5	12	1	42.1	-28.9	57.7
5-23	4728	194	63	21	3631	186	52	24.8	7.5	14.4 e)

a) Ordinal month 0 is the duration from 0 to 0.5 months. Ordinal month 1 is from duration 0.5 to 1.5 months. Ordinal month 2 is from duration 1.5 to 2.5 months, and so on.
b) Two times estimated standard deviation of estimated Swedish marriage rate, in percent of the latter.
c) If this had been 6, the bias would have been +9.4 percent, not -8.8 percent.
d) If this had been 3; the bias would only have been -4.3 percent, not -36.2 percent.
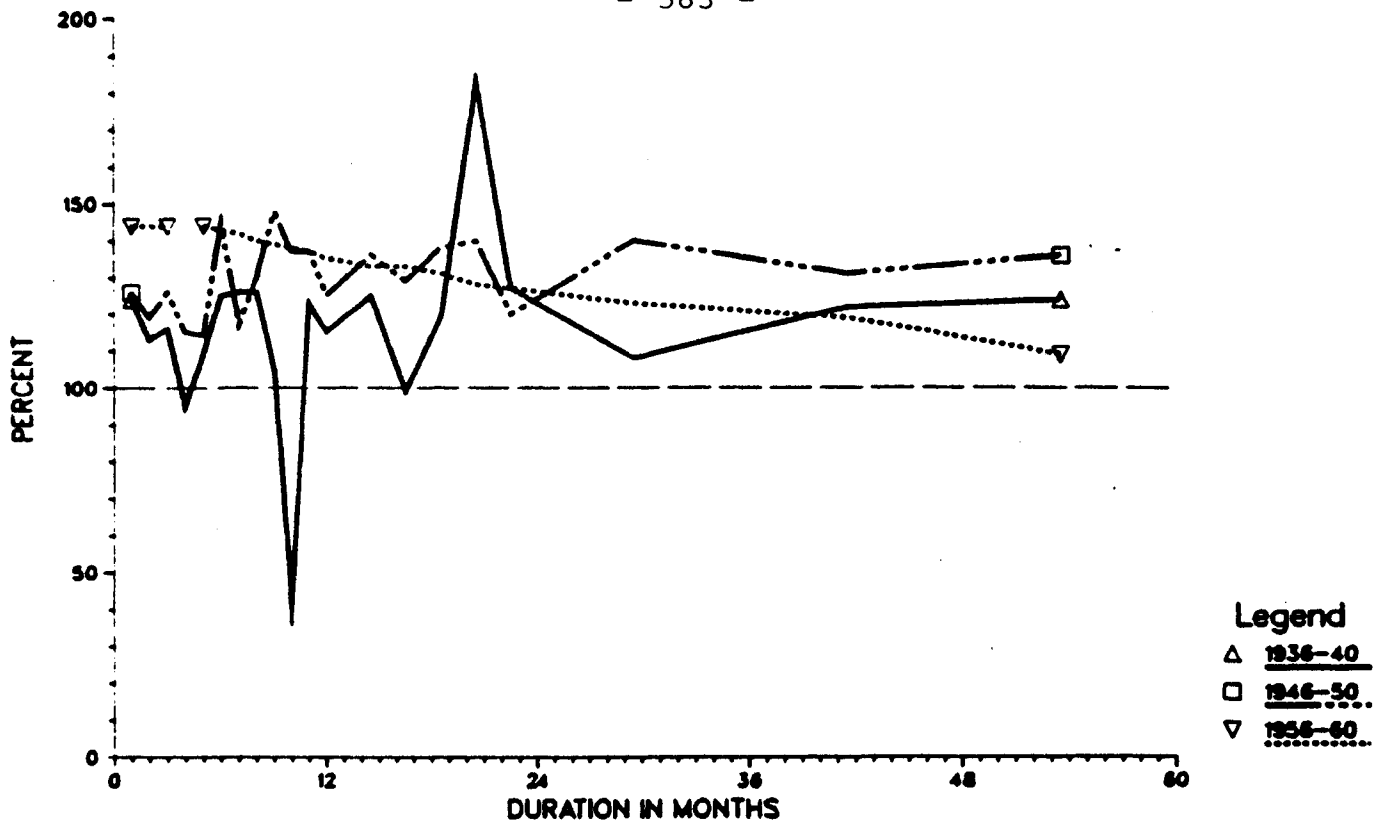e) Corresponding figure for first birth rate: 25.2.
*) Less than 10 marriages.

Figure 4. Distortion produced in marriage rates by applying
Danish observational plan to Swedish data. For
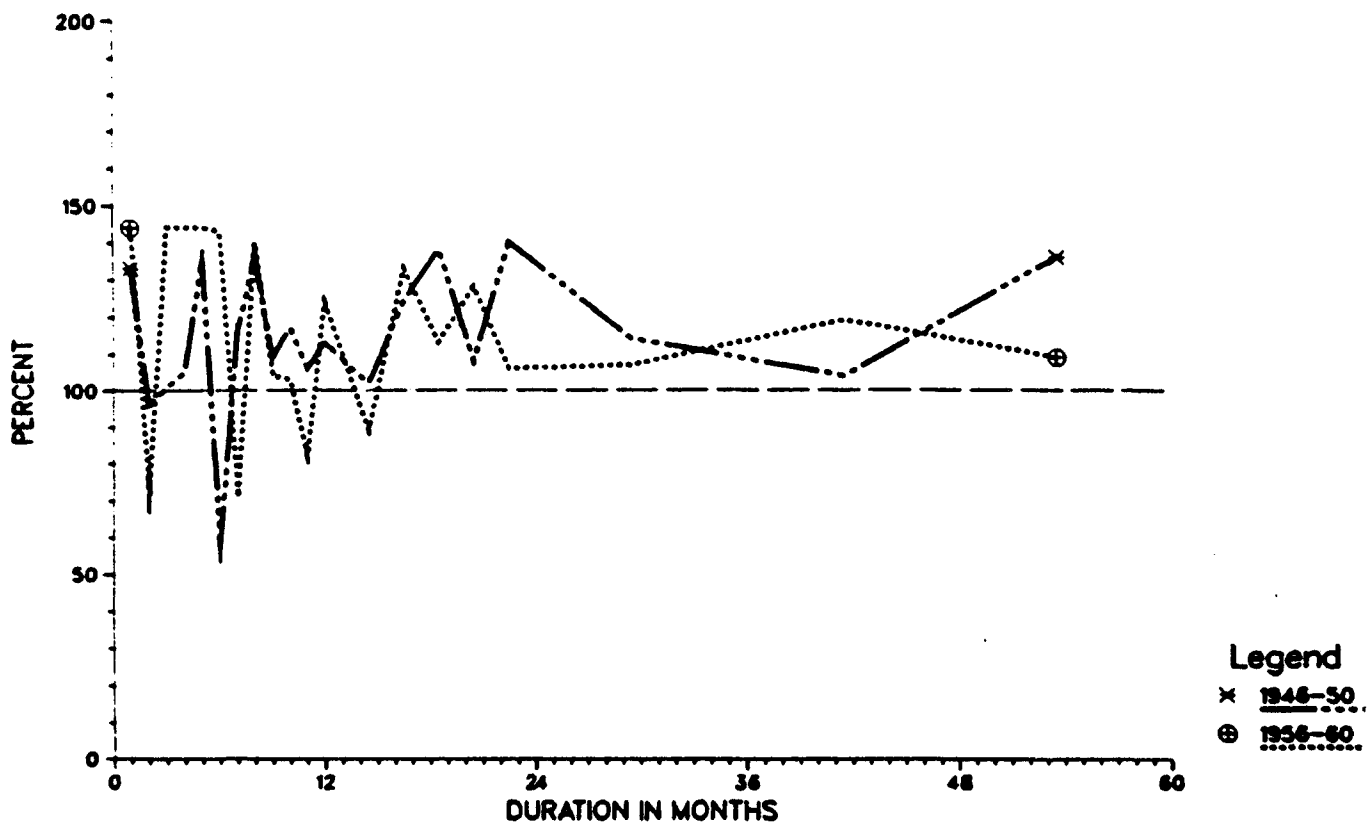selected cohorts, all starting ages combined.

Figure 5. Distortion produced in rates of first birth by applying
Danish observational plan to Swedish data. For selected
cohorts, all starting ages combined.

1. A negative bias (or equivalently, a distortion factor $1+\hat{b}$ less than 1) is possible for marriage rates as well as for first birth rates. We were alerted to this possibility for first birth rates by Hoem (1983), but the appearance of negative biases for marriage rates took us by surprise. It is a demonstration that misrecording of data for even only a few women caused by the observational plan (or otherwise) may have striking effects.

2. The vulnerability of bias estimates for short duration periods to random effects is demonstrated also by the peak at ordinal month 21 for the cohort of 1936-40 in Figure 4, as well as by the general raggedness of the bias sequences. The recording or nonrecording of single events may have strong effects, as demonstrated by Footnotes c and d in Table 6. For such reasons, we have combined the data for several ordinal months and have based our bias analysis on computations for longer intervals.

3. In spite of the nice monotonic appearance of, say, the bias curve for the cohort of 1956-60 in Figure 4, we have been unable to detect any systematic durational pattern in the biases, even after some careful minor grouping of the ordinal months. Therefore, our main bias analysis has been for the long interval of ordinal months 5 to 23 combined for marriage rates, and ordinal months 1 to 23 combined for first birth rates. The deletion of the first ordinal month(s) avoids initiation effects of a union. (For first birth rates, it also avoids some trivial problems mentioned in Endnote 3.) A separate analysis of the biases in first birth rates for ordinal months 5 to 23 combined gave no further insight and is not reported here. We have seen little point

in going beyond a cohabitational duration of two years with our data.

For cohabitational marriage rates of ordinal months 5 to 23 combined, Table 7 shows little pattern by cohort and starting age in the estimated biases due to the Danish observational design. It systematically over-estimates the tendency to marry, roughly by a fifth to a third (or some-times more). Nevertheless, the dramatic fall in marriage rates over co-horts is well represented, and so is the "overrisk" of marriage of women starting cohabitation at ages 20-24 as compared to those who start at ages below 20. A more compact impression of the trend in marriage formation over cohorts (and of the corresponding biases) results if starting ages below 20 and 20-24 are combined, as in Table 8.

For cohabitational first birth rates of ordinal months 1 to 23 com-bined, the biases are smaller than for corresponding marriage rates (Table 9)[4], and the birth rate biases are negative for our earliest co-hort. Otherwise, they tend to be around a fifth upwards for starting ages below 20, and roughly a tenth upwards for ages 20-24 at the start of cohabitation. This differential bias would tend to somewhat overrate the fertility differences by age at start of cohabitation. These smaller biases hardly mask the general stability in cohabitational first birth rates across cohorts or the differential by starting age. In fact, even our single outlier observation, namely the fall in the first birth rate for teenage Swedish women living in consensual unions in our youngest cohort, is picked up well by the biased procedure.

Our general conclusion of this analysis of the biases inherent in the Danish observational plan, therefore, is that the estimated "risk" levels may be somewhat biased, usually upwards, but not enough to cover up or mask any trends across cohorts or any differences by age at the start of the consensual union.

Table 7. Marriage rates for nulliparous Swedish women in consensual unions, at durations of 5 to 23 ordinal months combined. By cohort and by age at start of union. According to Swedish and Danish observational plan. Per 1000 women per month.

Cohort born in	below 20			Starting ages 20-24			25 and over			Proportionate increase below 20 to 20-24[e]	
	$S^{a)}$	$DK^{b)}$	Bias,$^{c)}$ percent	S	DK	Bias	S	DK	Bias	S	DK
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1936-40	56.7	64.4	15	68.6	80.1	17	45.6	52.0	14	21	24
1941-45	36.8	40.2	9	55.4	70.7	28	27.0	39.3	45	51	76
1946-50	18.8	25.7	37	26.3	33.3	27	19.1	27.4	44	40	30
1951-55	9.4	12.4	33	13.3	17.1	28	16.9	18.9	12	42	38
1956-60	5.5	8.0	45	9.0	10.5	17	n.d.$^{d)}$	n.d.	n.d.	64	31

a) Swedish observational plan.

b) Danish observational plan.

c) 100(S/DK-1).

d) No data.

e) (10)=100[(4)/(1)-1],
   (11)=100[(5)/(2)-1].

Table 8. Marriage rates for nulliparous Swedish and Danish
women in consensual unions at durations of 5 to 23
ordinal months combined, for all starting ages below
25 combined. By cohort. Per 1000 women per month.

Cohort born in	Swedish women			Danish women DK
	S[a]	DK[b]	Bias[c] (percent)	
	(1)	(2)	(3)	(4)
1936-40	62.5	73.3	17	80
1941-45	47.7	57.5	21	70
1946-50	23.8	31.0	30	50
1951-55	11.7	15.3	31	20
1956-60	6.5	8.8	36	n.d.[d]

Footnotes, see Table 7.

Table 9. First birth rates for Swedish women in consensual unions, at durations of 1 to 23 ordinal months combined. By cohort and by age at start of union. According to Swedish and Danish observational plan. Also comparable rates for Danish women. Per 1000 women per month.

Cohort born in	Starting ages below 20				Starting ages 20-24			
	Swedish women			Danish women, DK[b]	Swedish women			Danish women, DK
	S[a]	DK[b]	Bias[c] (percent)		S	DK	Bias	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1936-40	25.8	18.6	-28	5.3	17.3	13.1	-24	6.3
1941-45	22.5	20.8	-7	7.6	10.6	11.2	6	6.1
1946-50	20.9	25.6	27	10.6	9.6	10.6	10	2.8
1951-55	16.1	18.0	12	7.0	9.0	9.8	9	6.9
1956-60	9.8	11.8	20	n.d.[d]	12.6	14.0	11	n.d.
All cohorts[e]	15.9	16.7	5	8.2	10.1	10.9	8	5.2

a) - d) See Table 7.
e)   For Sweden: women born 1936-60.
        For Denmark: women born 1926-55.

This impression is reinforced when we juxtapose diagrams of schedules of duration-specific marriage or first birth rates to cohabiting women, as in Figures 6 and 7. Even though a detailed comparison of corresponding curves in each figure will reveal some distortion effects, these are really minor by comparison to the general impressions of duration dependence and cohort trends.
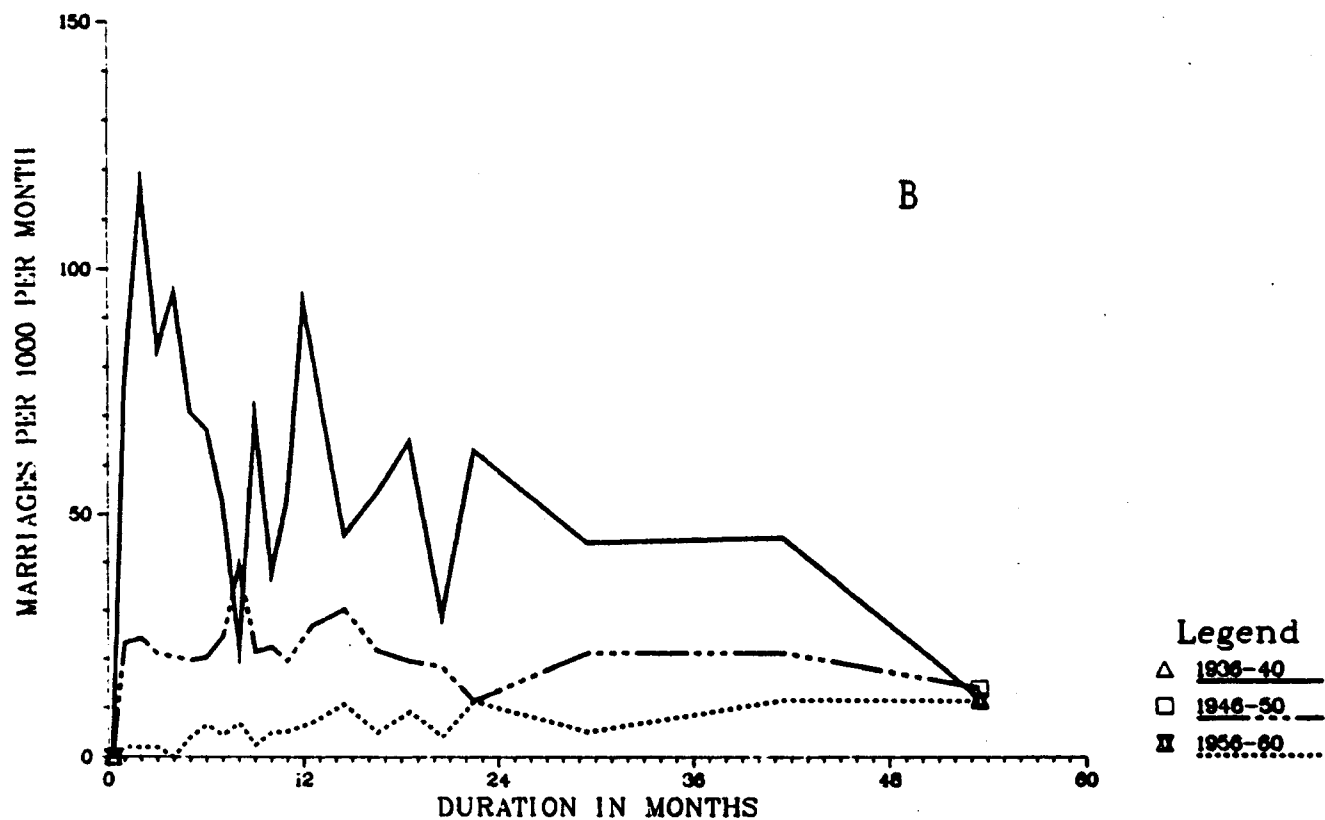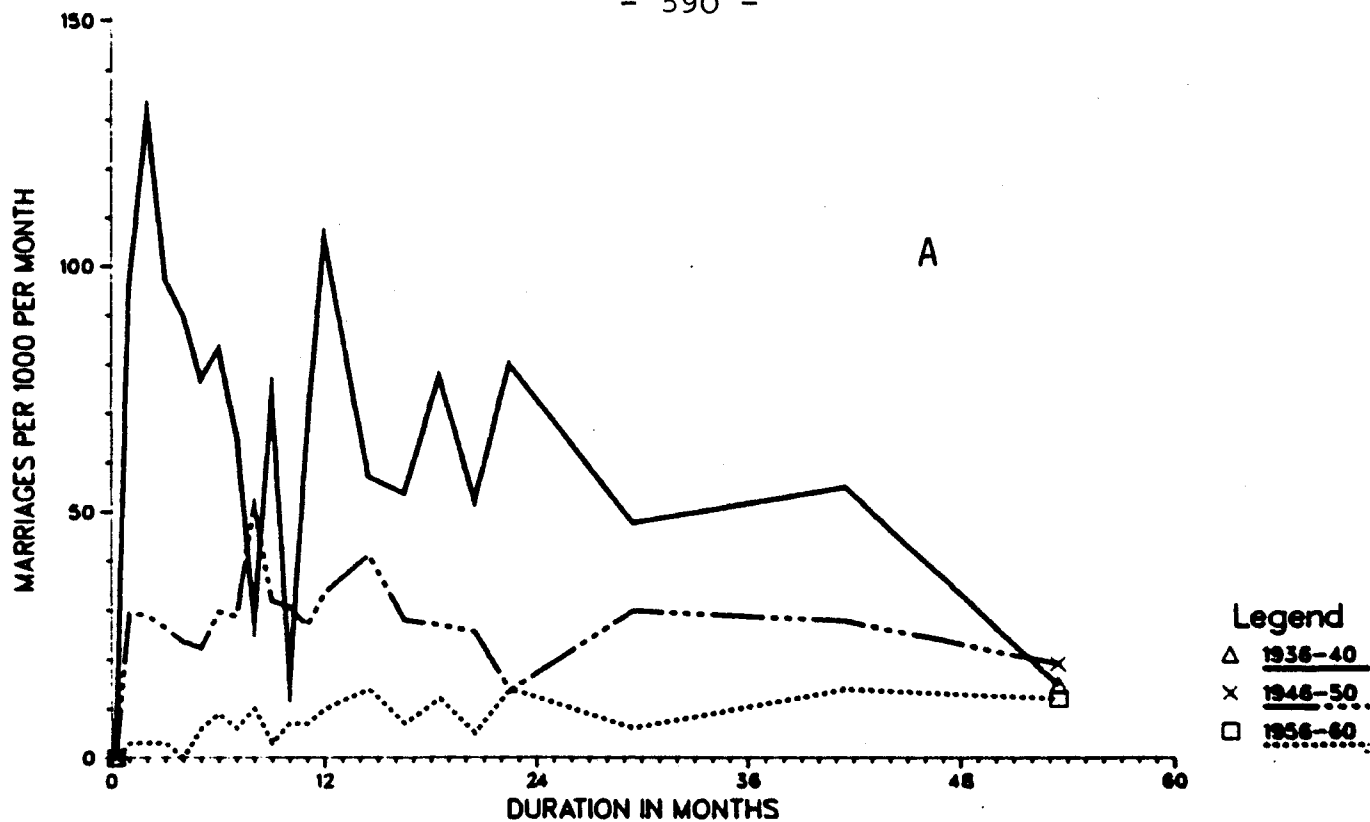
Figure 6. First marriages per 1000 cohabiting Swedish women at parity 0, by duration of consensual union. For selected cohorts, all starting ages combined. Panel A: According to imitated Danish design. Panel B: Swedish design.
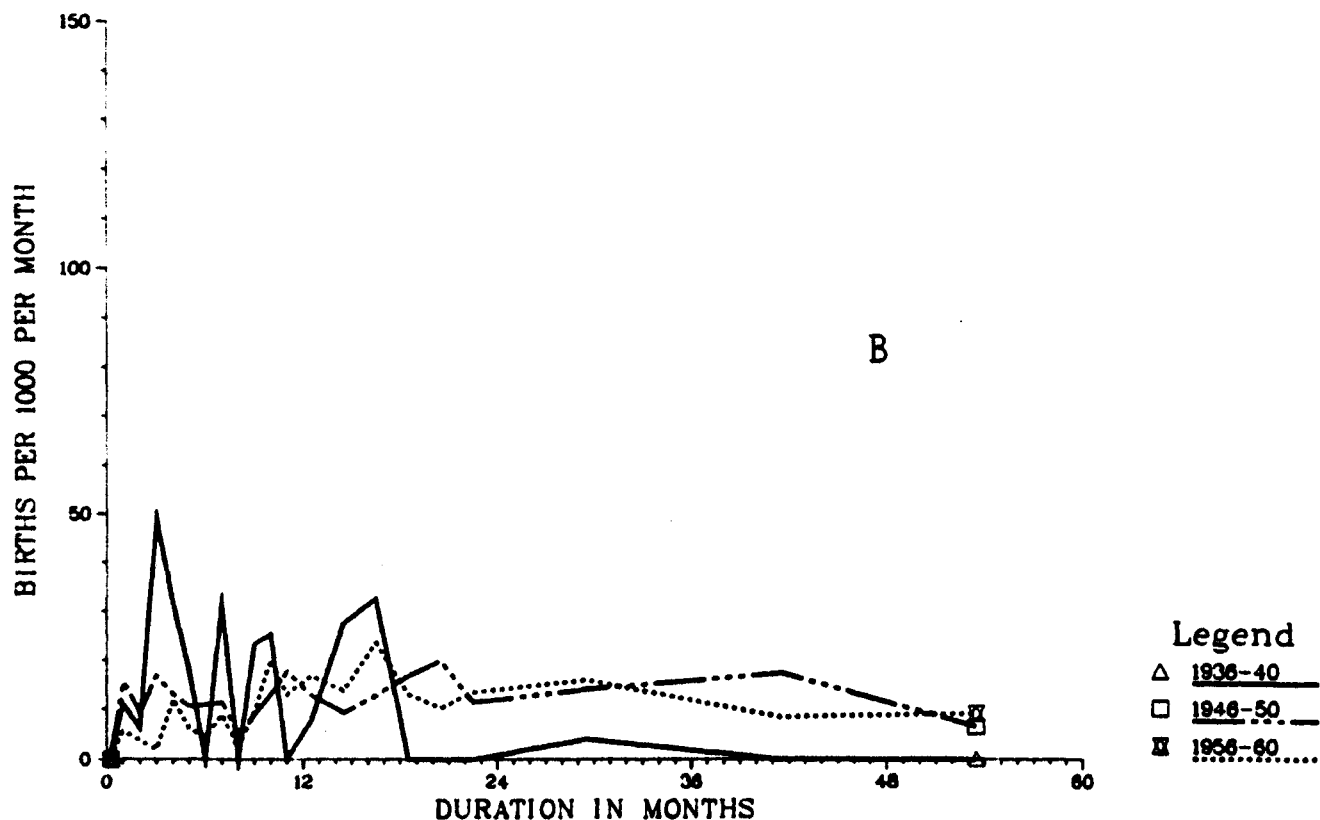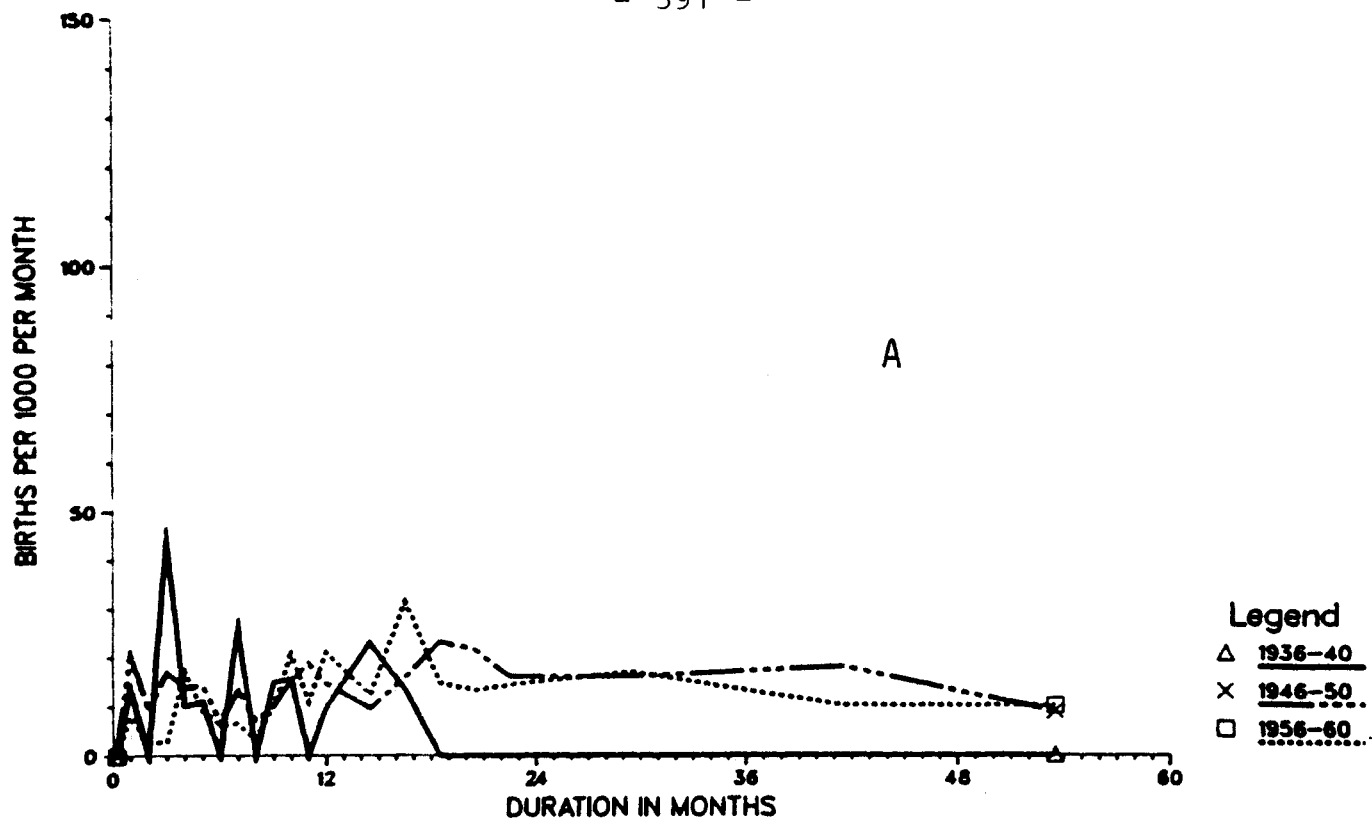
Figure 7. First births per 1000 cohabiting never-married Swedish women, by duration of consensual union. For selected cohorts, all starting ages combined. Panel A: According to imitated Danish design. Panel B: Swedish design.

4, A COMPARISON OF COHABITING NULLIPAROUS WOMEN IN SWEDEN AND DENMARK

Tables 8 and 9 invite a comparison between the behaviour in Denmark and Sweden. Table 8 (Colmuns 2 and 4) suggests that in each of the cohorts born in 1936 to 1955, nulliparous Danish women tended to marry to quite a larger extent than their Swedish counterparts. Even if we reduce the biased Danish marriage rates of the cohorts born in 1941-45 and 1946-50 by an amount intended to remove distortion[5] and bring them into line with the corresponding unbiased Swedish rates in Column 1, then the corresponding Danish partial[6] marriage probabilities over the 19 months in question are about 65 and 50 percent, respectively. Their Swedish counterparts are about 60 and 30 percent. For the cohort of 1951-55, the bias-adjusted Danish partial probability is 25 percent and the Swedish is 20 percent.

On the other hand, nulliparous Swedish women living in consensual unions have about twice as high first birth rates as do their Danish counterparts (Table 9). To see how this interacts with the differential nuptiality, we have also computed estimated partial probabilities $\hat{\pi}(z)$ of having a nonmarital first birth in a consensual union over $z = 19$ months when marriage is a competing risk but mortality and union dissolution are disregarded, as in Table 4, for Denmark and Sweden separately (Table 10). The Danish cohabitational marriage and first birth rates have been adjusted with the intention of removing design biases. In each column of Table 10, the first birth rate used is the same for each cohort since there are no great differences across cohorts, while the marriage rate is the one pertaining to each cohort. The marriage rate has been set to 0 in the final line of the table to get a "pure" fertility measure. The formula for $\pi(z)$ was given in connection with Table 4. A comparison between Tables 4 and 10 show the

negligible effect of the bias adjustment.

In each case, the childbearing probability for a Swedish woman is about twice its Danish counterpart, as were the birth rates themselves. By the standards of industrialized Western countries, Denmark probably has much nonmarital cohabitation (Brown and Kiernan, 1981) and a high incidence of cohabitational first births. By comparison to Sweden, however, it evidently was a laggard on both counts in the period we have investigated.

Table 10. Estimated probability $\hat{\pi}(z)$ of having a nonmarital first birth in a consensual union over a period of z=19 months under the competing risks of childbearing and marriage, disregarding mortality and union dissolution. By cohort and by age at start of cohabition. Complete Swedish data and adjusted Danish data. Percent.

Cohort born in	Starting ages below 20		Starting ages 20-24	
	Sweden[a]	Denmark[b]	Sweden[a]	Denmark[b]
	(1)	(2)	(3)	(4)
1936-40[c]	16.4	8.2	9.9	4.7
1941-45	19.0	8.3	11.0	5.4
1946-50	22.1	10.1	13.8	6.2
1951-55	23.9	12.0	15.5	7.5
1956-60	24.8	n.d.[d]	16.1	n.d.
No nuptiality	26.0	13.8	17.5	8.7

a) For Sweden: Marriage rates from Table 7, by cohort and starting age. First birth rates from Table 9, by starting age, for all cohorts born 1936-60 combined.

b) For Denmark: Marriage rates computed from Danish data and subsequently adjusted for design bias as described in Endnote 2, for each cohort and starting age separately, using items from Table 7 instead of Table 8 for the adjustment. First birth rates from Table 9, by starting age, for all cohorts born 1926-55 combined, reduced in each case by 0.4 per 1000 to adjust for design bias.

c) For Denmark: 1926-40, combined.

d) No data.

ACKNOWLEDGEMENTS

ENDNOTES

1. Some mathematics explaining this feature can be found in Hoem (1985, Section 2.3). Allison (1985) recently reviewed items concerning the closely related notion of backward recurrence times.

2. The use of information from populations different from but similar to the study population to supplement incomplete data from the latter is commonplace in demography. Indeed, the whole notion of model life tables and much of the practical use of stable population theory are based on this idea.

3. This makes ordinal month 0 last, on the average, from an "exact" cohabitational duration of 0 to 1 month, it makes ordinal month 1 last between durations 1 and 2 months in continuous time, and so on. If all events were to be recorded as happening in the middle of the calendar month of reported occurrence, ordinal month 0 should actually have been treated as lasting between "exact" durations 0 and 0.5 months, ordinal month 1 as being from 0.5 to 1.5 months, and so on. The effect would be to cut the exposures for ordinal month 0 in half and thus double the occurrence/exposure rates of this month. The rates for the other months would have remained unchanged. We only thought of this after all of our analysis of the Danish data had been completed (but before we analyzed the Swedish data), and it has not been worthwhile to go back and make the trivial corrections which such increased accuracy would lead to in a small detail of our results.

The treatment of the events of ordinal month 0 also leads to a second problem of the same minor importance for our analysis. If you are told that a woman both started a consensual union and had her first

birth in April, 1977, you do not know whether she had the baby first (in which case she would not be eligible for inclusion in our analysis of the behavior of nulliparous unmarried cohabiting women) or last (in which case she would be eligible). The way questions were phrased in the Danish questionnaire, we do know whether marriage or cohabitation started first if both were recorded in the same calendar month and year for a woman. For simplicity, therefore, the data were treated as if start of cohabitation always preceded any other event in the same calendar month.

In the Swedish survey, women were not supposed to report cohabitations lasting less than a month. They were asked to state the calendar month in which they moved together with their man for each union, marital as well as consensual. We have recorded a union as consensual, therefore, if and only if the reported start of the union preceded the month of marriage (if any) by at least on calendar month. By analogy, births have been treated as starting before any union reported in the same calendar month. Unions with a reported dissolution in the same month as the start would be ignored. In effect, therefore, no event can be recorded in ordinal month 0 of a union in the Swedish data.

In our comparisons of our Danish and Swedish data, the two sets have been brought into line by the trivial remedy of ignoring events in ordinal month 0 in the Danish data.

4. The rates of Columns 4 and 8 in Table 9 differ somewhat from orresponding rates in Table 3 because they have been computed for a lightly different set of ordinal months of duration.

5. Divide the rate in Column 4 of Table 8 by the item in Column 2 and multiply by the item in Column 1. Then the bias in the Danish rate will be removed, provided the proportionate biases are the same in Denmark and Sweden.

6. As in Column 7 of Table 2, where corresponding unadjusted Danish partial probabilities were listed, the figures given here are estimates of the single decrement life table probability of marrying at some time during the 19 months in the absence of any competing risks (death, union dissolution, first birth). It is called a partial probability because one only accounts for part of (some of) the risks really involved.

REFERENCES

Allison, Paul D. (1985). Survival analysis of backward recurrence times. Journal of the American Statistical Association 80, 315-322.

Arvidsson, Arne et al. 1980. Kvinnor och barn: Intervjuer med kvinnor om familj och arbete. Stockholm: National Central Bureau of Statistics, Information i prognosfrågor 1982:4.

Brown, Audrey and Kathleen Kiernan (1981). Cohabitation in Great Britain: Evidence from the General Household Survey. Population Trends 25(1), 4-10.

Finnäs, Fjalar. 1983. Fertility trends of Swedish women, 1960-1977: A study of birth intervals in selected cohorts of women born between 1927 and 1952. Pp. 133-217 in Leif Johansson and Fjalar Finnäs, Fertility of Swedish women born 1927-1960. Stockholm: Statistics Sweden, Urval No. 14.

Finnäs, Fjalar and Jan M. Hoem. 1980. Starting age and subsequent birth intervals in cohabitational unions in current Danish cohorts, 1975. Demography 17:275-295.

Hobcraft, J.N. and G. Rodriguez. 1980. Methodological issues in life table analysis of life histories. Paper prepared for the IUSSP-WFS-CPS Seminar on the Analysis of Maternity Histories, London, April 1980.

Hoem, Jan M. 1983. Distortions caused by nonobservation of periods of cohabitation before the latest. Demography 20(4), 491-506.

Hoem, Jan M. (1985). Weighting, misclassification, and other issues in the analysis of survey samples of life histories. Pp. 249-293 in J.J. Heckman and Burton Singer, Longitudinal Analysis of Labor Market Data. Cambridge University Press.

Hoem, Jan M. and Bo Rennermalm (1985). Modern family initiation in Sweden: Experience of women born between 1936 and 1960. European Journal of Population 1(1), 81-112.

Hoem, Jan M. and Randi Selmer. 1984. The negligible influence of pre-marital cohabitation on marital fertility in current Danish cohorts, 1975. Demography 21(2), 193-206.

Hyrenius, Hannes (1941). Studier rörande den utomäktenskapliga fruktsamhetens variationer. Skrift no. 28 utgiven av Fahlbeckska stiftelsen. Lund (Sweden): Gleerup.

Lyberg, Ingrid (1984). Att fråga om barn: Teknisk beskrivning av undersökningen "Kvinnor i Sverige". Stockholm: Statistics Sweden. Bakgrundsmaterial från Prognosinstitutet, 1984:4.

Matović, Margareta R. (1985). "The Stockholm Marriage": Family Formation and Choice of Partner in Stockholm 1850-1890. In Swedish, with a long summary in English. Stockholm: Liber Förlag.

Page, H.J.; B. Ferry; J.H. Shah; and R. Lesthaeghe. 1980. The most recent births: Analytical potential and some underlying problems. Paper prepared for the IUSSP-WFS-CPS Seminar on the Analysis of Maternity Histories, London, April 1980.

Qvist, Jan and Bo Rennermalm. 1985. Att bilda familj: Samboende, äktenskap och barnafödande bland kvinnor födda 1936-60. Stockholm: Statistics Sweden, Urval No. 17.

Rindfuss, Ronald R.; Larry L. Bumpass; and James A. Palmore (1985). Analyzing selected fertility histories: Does the restriction bias the results. Unpublished manuscript.

Selmer, Randi. 1984. Cohabitation without marriage - marriage and births [in Norwegian cohorts]. (In Norwegian with a summary in English.) Oslo: Central Bureau of Statistics of Norway, Artikler No. 146.

Trost, Jan. 1978. A renewed social institution: Non-marital cohabitation. Acta Sociologica 21(4), 303-315.

Compound arrival times

Peter Mitter

The basic model of survival analysis corresponds to a stochastic process $\{X_t\}_{t \geq 0}$ with two states only: a transient origin state S_o, $Pr(X_o = S_o) = 1$ and an absorbing destination state S_d, arrival time being time before absorption.

In many social science applications this model is not adequate, because there are several transient states (one of them being the origin state), and possibly also several absorbing states or absorbing sets of states, though (observed) arrival time is still time before absorption. Birth intervals of women e.g. are composed of (i) the period of post partum amenorrhea following the first birth (ii) the total duration of menstruating intervals between the two births (iii) the period of pregnancy associated with the latter birth. If there are abortions or still births between the two live births, then the corresponding periods of pregnancy and post termination amenorrhea are also part of the birth interval (Pathak and Sastry, 1983).

Almost all applications where arrival time is age at some event concern two periods at least: one period until maturity with respect to the event under consideration is reached, and a subsequent period where the individual is exposed to the corresponding risk. Age at first marriage may thus be decomposed into the phase until reaching marriageability and the subsequent phase until marriage (whereby the latter phase may additionally be decomposed in suitable sub-phases, see e.g. Coale and McNeil, 1972).

In contrast to these two examples, it moreover may happen that the transient states are not necessarily passed in a predetermined order. Imagine a system where

transition is possible between any two of several unobserved - maybe latent - states, but where again only total time until absorption can be observed. Cohen (1963) and Rapoport and Chammah (1965) consider Markov chain models with similar properties (Cohen in an analysis of Asch type experiments on conformity pressures, Rapoport and Chammah in their analysis of iterated Prisoner's Dilemma experiments). The extension of such models to continuous time survival models is obvious (see e.g. Bartholomew, 1973: 191ff, for a model of job duration).

There are two arguments in favor of an explicit modelling of compound arrival times. The first argument is a technical one. If the composition character is neglected, i.e. if different transient states are treated as if they were equivalent, then the stochastic properties of the process are altered: in general they become more complicated, although the state space is reduced. If e.g. the states of a Markov process are aggregated, then (apart from special cases) the resulting process will not be Markovian anymore. As a result one gets a problem of mis-specification. The second argument is still more important. In social sciences it is not only important to fit observed distributions as good as possible, but to explain them in terms of their generating processes. Consider again age at first marriage. It is good to find a model which fits the corresponding age distribution fairly well. But is better to have a model which additionally allows to estimate the distribution of age when marriageability is reached.

Assume now the simplest case of arrival times which are sums of independent components. As is well-known, the distribution of such a sum is given by the convolution of the component distributions, i.e. if X and Y are independent continuous random variables with d.f. f_X, f_Y, and p.d.f. F_X, F_Y, respectively, then $Z = X + Y$ has density function

$$f_X * f_Y(z) = \int_{-\infty}^{\infty} f_X(s) \, f_Y(z-s) \, ds = \int_{-\infty}^{\infty} f_X(z-s) \, f_Y(s) \, ds$$

and probability distribution function

$$F_X * F_Y(z) = \int_{-\infty}^{\infty} F_X(z-s) \, f_Y(s) \, ds = \int_{-\infty}^{\infty} F_Y(z-s) \, f_X(s) \, ds$$

There are families of distributions which are closed under the convolution operation. The convolution of two Normal distributions is again Normal, the convolution of two Cauchy distributions is again Cauchy, the convolution of two Gamma distributions with fixed scale parameter is again Gamma distributed. As a consequence, there is no chance to identify the components of a convolution, if

they belong to the same parametric family like their sum.

A second problem concerns properties which are not preserved under the convolution operation. In contrast to intuition, the convolution of two unimodal distributions is not necessarily unimodal (a counterexample is given by Feller, 1966: 164). This is not a very pleasant possibility for situations where theoretical distributions are expected to be unimodal. But life is not as harsh. Our most familiar distribution functions like the Normal, the Exponential, the Weibull (with $p \gtrless 1$) and the Gamma distribution (with $p \gtrless 1$) are strongly unimodal. This means, that any convolution of them with an arbitrary unimodal distribution is again unimodal, and convolutions of strongly unimodal distributions are even strongly unimodal (Medgyessy, 1977: 24ff, where also conditions for strong unimodality are given and families of unimodal distributions closed under convolution are studied). Thus if a compound arrival time is modelled as the sum of independent components of one of the types given above, plus possibly one additional component which is unimodal, but not necessarily strongly unimodal, then the "theoretical" distribution of the sum is unimodal.

This paper deals with convolutions where one of the components is Normal. One shortcoming of all commonly used survival time distribution models is that they do not have a location parameter. Thus a shift of the time origin without changing the shape of the distribution is only possible by the introduction of an additional parameter. Modelling a convolution with a Normal component not only allows for the introduction of such a shift (e.g. age when maturity is reached), but also for heterogeneity with respect to this shift (e.g. differences in maturity). Consider the case of the convolution of a Normal $(N(\mu, \sigma^2))$ and an Exponential (λ) distribution. The corresponding integrals are not solvable in terms of elementary functions, but suitable transformations yield the following expressions for density f and distribution function F of the convolution:

$$f(x) = \lambda e^{-\lambda(x-(\mu+\lambda\sigma^2/2))} \Phi(\frac{x-(\mu+\lambda\sigma^2)}{\sigma})$$

$$F(x) = \Phi(\frac{x-\mu}{\sigma}) - e^{-\lambda(x-(\mu+\lambda\sigma^2/2))} \Phi(\frac{x-(\mu+\lambda\sigma^2)}{\sigma})$$

where Φ is the standard Normal distribution function. As most mathematical subroutine libraries contain procedures to calculate ϕ, there should be no problems with the calculation of f and F, or, as a consequence, with maximum likelihood estimation. The only problem which may arise is the following. Both f and F

contain functions of the type $e^{-\lambda\sigma z}\phi(z)$ with $z = (x-(\mu+\lambda\sigma^2))/\sigma$. If λ or σ (or both) are large, then, in the most relevant domain around the mean $\mu + 1/\lambda$, the computation of $f(z)$ and $F(z)$ requires the multiplication of a very large number $e^{-\lambda\sigma z}$ with a very small number $\phi(z)$ which easily ends in arithmetic overflow or numerical inaccuracies. Fortunately, this is the case when the convolution model is not very appropriate. If λ is large, then the exponential arrival times are short, and the compound distribution will not seriously deviate from its Normal component. If σ is large, then there is not much to be concluded beyond the fact that there is much unexplained variance.

In the case of a Normal and more than one exponential delays two cases should be distinguished

(i)   all exponential delays are identically (but independently) distributed. Then their sum is Gamma-distributed. There is no manageable representation of the sum of a Normal and a Gamma distribution. On the other side, this does not seem to be a generic case. Apart from special situations (an example is given later in this paper) there is no general plausible reason why sojourn times spent in adjacent, distinct phases should be identical distributed. The even more complicated case of some delays which are identical distributed while others are not is not considered because of the same reason.

(ii)   all exponential delays stem from different (and independent) distributions (i.e. distributions with different λ). Then the distribution of their sum is a linear combination of the component distributions (see e.g. Feller, 1966: 40). The convolution of this sum with a Normal is then the corresponding linear combination of the convolutions of the Normal distribution with the respective exponential delays. In comparison to the bivariate case studied above there are no additional technical difficulties. Having in mind the computational problems mentioned above one should nevertheless be aware that with more and more delays taken into the model the corresponding λ estimates will become larger and larger. Note that in the linear combination mentioned above there are also negative summands, it is thus not possible to misinterpret this convolution as a simple mixture of Exponential distributions.

Let us finally examine age as first marriage of women as an example of a compound arrival time. This problem was considered by Coale and McNeil (1972), but they put the cart before the horse. They show that the corresponding age distribution can be approximated very closely by the limiting distribution of the

convolution of an infinite number of (normalized) exponentially distributed components, and that an equally good fit is obtained by the convolution of a finite number of the exponential components plus an additional Normal distribution. The parameters of the exponential components are, however, generated by the simple formula $\lambda_n = a + n.b$ which is not confirmed by the authors but because of computational reasons (the specification reminds of Polya's model of infection).

As an alternative, consider the age of entry into a marriageable state as a variate which is not further explained but assumed to be Normally distributed. The subsequent phase until marriage may then be interpreted as a sequence of sub-phases such that the begin of each sub-phase coincides with beginning to date a new partner which may become a spouse (the last sub-phase is the period of dating the future spouse and the period of engagement). Each dating may additionally be interpreted as a Bernoulli trial leading to marriage (with some probability q) or to beginning to date with a new partner (with probability p = 1-q). To make things easier, assume that all dating periods (including the last one) are i.i.d., the common distribution being exponential with parameter λ, and that the "success probability" p is also equal for all periods. Then the time between entry into marriageability and marriage is a mixture of Gamma distributions with geometric weights

$$f(z) = \sum_{n=0}^{\infty} q \, p^n \, \lambda \, \frac{(\lambda z)^n}{n!} \, e^{-\lambda z} = \lambda \, q \, e^{-\lambda q z}$$

thus again exponentially distributed with parameter λq. The distribution of age at first marriage is then the convolution of a Normal and an Exponential distribution. Note that in this model the parameters λ and q cannot be identified from the survival time distribution (the product λq can, of course).
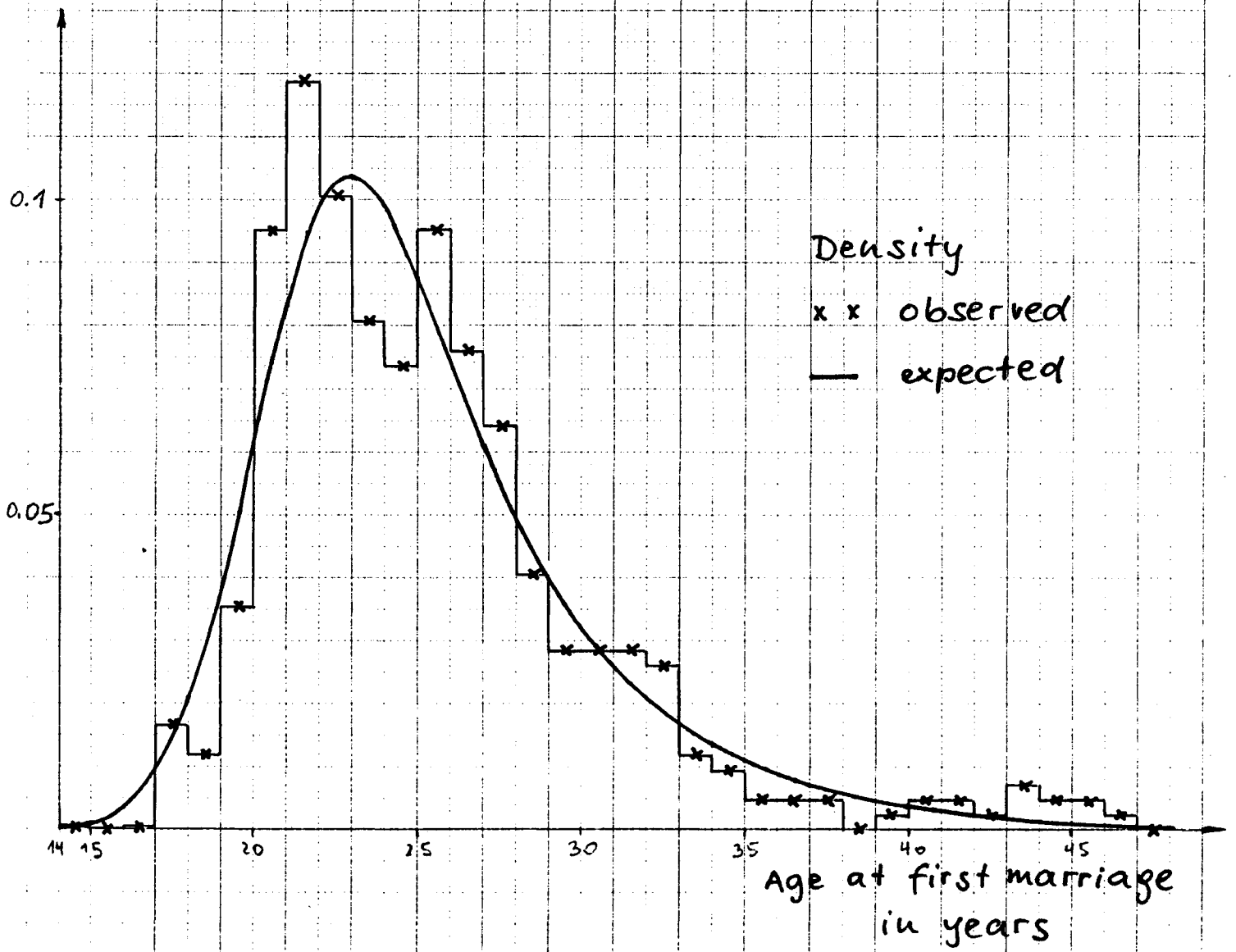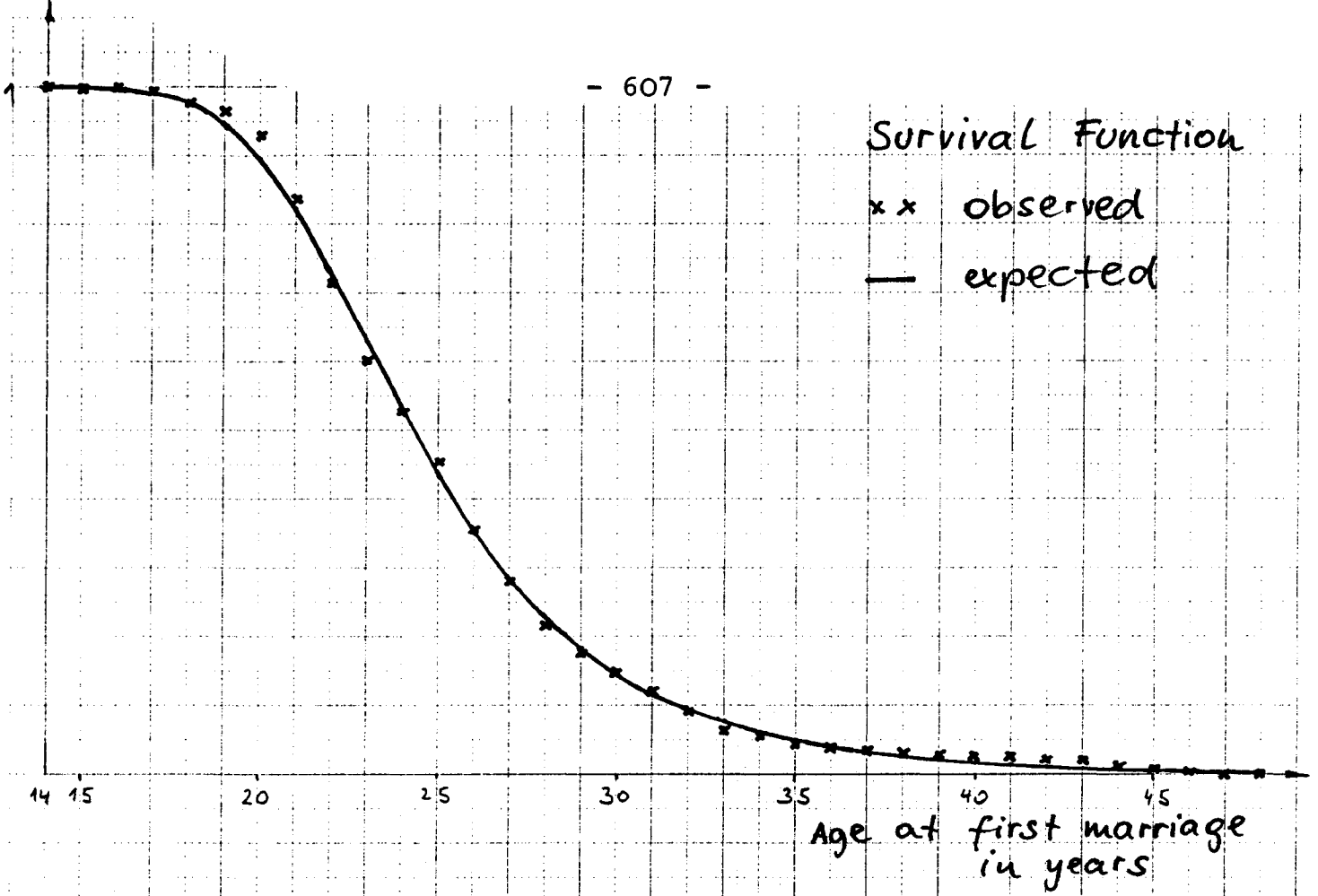
The parameters μ, σ and $\alpha = \lambda q$ of this models are estimated from a birth cohort of 421 Austrian women born in 1926 (source: Mikrozensus 1976, the cohort has reached age 50 at the time of interview, a time span which is assumed to be long enough to cover the event "first marriage", thus the sample includes only women which are or have been married). Both birth and first marriage date were recorded by month. Instead of maximum likelihood the method of moments was used, because the respective equations are very simple:

mean	$m = \mu + 1/\lambda$
variance	$s^2 = \sigma^2 + 1/\lambda^2$
3rd c.m.	$\mu_3 = 0 + 2/\lambda^3$

They lead to an estimated mean age at entry into marraigeability of 20.6 years (with standard deviation of 2.3 years) and a subsequent mean premarriage time of 4.6 years. If the efficiency of the moment estimator is not supposed to be acceptable, these values can be used as starting values for MLE.

In the two figures observed and expected age distributions are compared ("observed" values are life table estimates, grouping of age in years). There is fairly good coincidence not only between the p.d.f.'s (where one generally expects some similarity), but also between the two densities. A refinement fo this model as a convolution of a Normal plus two exponential delays is closer at Coale and McNeil's study. It should result in an even better fit - of course. In the light of the explanation above this could be generated by a different distribution in the last sub-phase (dating with future spouse and engagement). But one should be careful: there is no chance to decide which exponential belongs to which period (because of a fundamental identification problem due to the commutativity of sums). The conclusion that the last phase has the shortest average duration may be valid on plausible grounds, but it cannot be infered with statistical methods.

To sum up, it is very advisable to model composed arrival times because of two reasons. First, we may obtain distribution families which are both more applicable than our usual familiar spectrum of functions and more relevant with respect to the process under study. And second, there are good chances that we obtain additional informations about the process under study, despite of the identification problems mentioned above.

Survival Function
x x observed
— expected

Age at first marriage
in years

Density
x x observed
— expected

Age at first marriage
in years

REFERENCES

Bartholomew, D.J. (1973), Stochastic Models for Social Processes, 2nd ed., London: Wiley.

Coale, A.J. and D.R. McNeil (1972), The Distribution by Age of the Frequency of First Marriage in a Female Cohort, Journal of the American Statistical Association, Vol. 67, No. 340: 743-749.

Cohen, B.P. (1963), Conflict and Conformity, Cambridge, Mass.: M.I.T. Press.

Feller, W.V. (1966), An Introduction to Probability Theory and Its Applications, Volume II, New York: Wiley.

Medgyessy, P. (1977), Decomposition of Superpositions of Density Functions and Discrete Distributions, Bristol: Adam Hilger.

Pathak, K.B., and V.S. Sastry (1983), A Modified Stochastic Model for Closed Birth Interval, Journal of Mathematical Sociology Vol. 9: 155-163.

Rapoport, A. and A.M. Chammah (1965), Prisoner's Dilemma. Ann Arbor: University of Michigan Press.

Regression analysis for discrete event history or failure time data

Alfred Hamerle

1. Introduction

Event history or failure time data are collected in follow-up studies, retrospective studies, and sometimes in longitudinal panels. The data record qualitative changes over time in some important variables. The main purpose of the statistical analysis of such event histories or failure times is to investigate the time it takes before a certain event occurs. Examples are job changes, changes in residence, lay-offs, births, marriages, divorces, deaths, etc. In addition, it is important to evaluate the association of exposure, treatment and prognostic factors with the distribution of time until the event occurs.

Sometimes there is only one episode or spell for each individual measuring the time interval between an initial event and a termination event. This applies in particular to survival analysis where the detection of a disease is the initial event and the patient's death is the termination event. In other applications of these methods individuals can experience repeatable events or failures and moreover, these events or failures may be of various kinds. This leads to general multiepisode - multistate

models and the successive episodes represent durations in different states.

For the statistical analysis of such dynamic processes hazard rate models can be used where the hazard rate depends on independent variables. The statistical theory of duration data using hazard rate models is described by Kalbfleisch and Prentice (1980), Lawless (1982), Cox and Oakes (1984), Tuma and Hannan (1984), Blossfeld, Hamerle and Mayer (1986) and others. Hamerle (1984) surveys applications of duration models in different areas and considers a general approach in the multistate - multiepisode case. Hujer and Schneider (1986) investigate the data of the first wave of the Socioeconomic Panel and compare several hazard rate specifications.

Most of the methods for analyzing event history or failure time data assume that time is measured as a continuous variable. The analysis pre-sented here is specifically intended for situations in which the time scale is genuinely discrete or in which there is substantial grouping of the response times into class intervals. The methods are applicable when the data are not available as essentially exact response times but when the data record only the particular interval of time in which each event or failure occurs. This applies in particular to longitudinal panels where event histories e.g. about employment status and other important qualitative changes between the successive panel waves are registered retrospectively in fixed-length periods. One of the new panel studies of this kind is the Socioeconomic Panel of the 'Sonderforschungsbereich 3" (see Hanefeld (1984)). Other applications are in medical work when patients are followed up and detailed information on each patient is collected at fixed intervals, or in sociological research when attention is given to qualitative changes that occur in specific time intervals.

If there are only a few time intervals or if the time units are large then many failures are reported at the same time and the number of ties becomes high. Then, strictly speaking, continuous-time techniques are inappropriate (Cox and Oakes, 1984, p. 99/100). Some continuous-time methods, especially the partial likelihood estimation procedure for Cox's Proportional Hazards model (see e.g. Kalbfleisch and Prentice, 1980, ch. 4) make use of the temporal order in which the failures or events occur and they cannot be applied directly when the data include tied observations. In the presence of ties an approximate partial likeli-hood function is widely used (Breslow, 1974). But when the number of

ties becomes high, this approximation yields severely biased estimates (Cox and Oakes, 1984, p.103, Kalbfleisch and Prentice, 1980, p. 74/75). Kalbfleisch and Prentice (1980), p.75, emphasize that there is some asymptotic bias in both the estimation of the regression coefficients and in the estimation of its covariance matrix. This applies not only to the Cox model but also to fully parametrized specifications of the hazard rate. Moreover, the papers which deal with the derivation of the asymptotic properties of the estimators in hazard rate models (see e.g. Andersen and Gill (1982), Borgan (1984)) assume that ties only occur with zero probability.

In such situations discrete-time models are more suited for the analysis of failure time data. Several authors, including Thompson (1977), Prentice and Gloeckler (1978), Mantel and Hankey (1978), Allison (1982), Aranda-Ordaz (1983), Laird and Olivier (1981), Hamerle (1985), have studied discrete-time regression models for failure time data. Here we present a different approach and consider the general case where individuals can experience many events or failures as time goes on and the events or failures may be of multiple kinds. In addition to the failure times and types of failures of an individual some concomittant information on explanatory variables or prognostic factors is included in the model to study the relationships between these variables and failures. The covariates may be time dependent and random with distributions depending on the observed experimental history. The covariate process may include fixed and 'external' time dependent as well as 'internal' time dependent covariates (see Kalbfleisch and Prentice, 1980, ch.5.2). The models also contain individual specific parameters. The role of the individual specific components is to control for unobserved heterogeneity, e.g. for omitted variables. The regression coefficients which are common to all of the individuals in the sample may be time varying, i.e. they may depend on the time intervals. Here we present several discrete-time hazard rate regression models and discuss their advantages and limitations. We derive unconditional, conditional and marginal estimation procedures where our concern is with the regression parameters β_t, the structural parameters of the model.

We refer to related work by Heckman (1981a, 1981b), Chamberlain (1980,1985), Arjas (1984) and Arjas and Haara (1986). Heckman and Chamberlain consider discrete-time models for state probabilities in analyzing traditional

panel data. With the exception of Chamberlain (1980) these studies only investigate the case of two states. The models are appropriate for typical panel studies where the individual's state is determined at particular points in time and where information about events between these successive points in time is not available. Here we are modeling transition probabilities which is more general if the state space contains more than two elements. But appropriate data is needed. 'Panel mortality' can also be incorporated. Our approach differs from the one considered by Arjas and Haara in that they use a binary logistic model and do not include individual specific parameters in their model. They derive asymptotic results for the estimated regression coefficients as the number of time intervals tends to infinity. Here we assume that observation time is finite and that there is a reasonable number of study subjects. Asymptotic properties here always concern the case where the number of study subjects tends to infinity.

2. A general discrete-time hazard rate model

Choosing some convenient point in real time as the origin, we split the time axis into successive intervals $t=1,2,\ldots$ The last time interval of the observation period is denoted by T and we consider probability models in discrete time $t=1,\ldots,T$.

The individuals included in the study are indexed by i, $i \geq 1$. It is not necessary for all the individuals to be present at the beginning of the observation period. Some of the individuals may join the study as time goes on. Let $z_i(t) \in \{1,\ldots,J\}$ denote the state in which individual i is at the beginning of time interval t. We define the indicator variables

$$Y_{ij}(t) = \begin{cases} 1 & \text{if individual } i \text{ is at risk at the beginning of} \\ & \text{time interval } t \text{ and } z_i(t)=j \\ 0 & \text{otherwise} \end{cases} \quad (2.1)$$

The individuals with $Y_{ij}(t)=1$ constitute the risk sets

$$R_j(t) = \{i: Y_{ij}(t)=1\} , \qquad j=1,\ldots,J. \quad (2.2)$$

$R_j(t)$ contains all individuals who are at risk during time interval t and who are in state j. Note that an individual cannot belong to more than one risk set in time interval t. If the sample size n is fixed in advance, then $i=1,\ldots,n$. If new individuals join the study as time goes

on, let n_t denote the number of individuals in the study up to time interval t. Then individuals i=1,...,n_t will have to be investigated to determine the risk sets $R_j(t)$.

Suppose then, that for every individual i and time interval t such that $Y_{ij}(t)=1$ for some j∈{1,...,J}, a p-vector $X_i(t)$ of relevant covariates is measured. The covariate process may include fixed or external time dependent as well as internal time dependent covariates (see Kalbfleisch and Prentice, 1980, ch.5.2). The vector of covariates may contain metric or dummy variables or both. As an approximation we assume that an individual can experience at most one event in time interval t, and if an individual is censored or is lost from the study, it is assumed that this happens at the end of the time interval. Similarly we think of the covariates as remaining fixed during each time interval, with the possible new value always determined at the beginning of the interval. When the interval lengths are small, this approximation is unlikely to influence statistical analysis a great deal.

Let $M(z_i(t))$ denote the set of attainable states from state $z_i(t)$ ordered in some way. Then we define the random variables $D_{ik}(t)$ as follows

$$D_{ik}(t) = \begin{cases} 1 & \text{if individual i moves to state k in time interval t} \\ 0 & \text{otherwise,} \qquad k \in M(z_i(t)). \end{cases} \qquad (2.3)$$

Furthermore, let $D_i(t)$ be the vector variable $\{D_{ik}(t): k \in M(z_i(t))\}$.

The history of the process up to time interval t is given by

$$F_t = \{(R_j(s), (X_i(s), D_i(s), i \in R_j(s)), j=1,...,J), s \leq t\}$$

and (2.4)

$$G_{tj} = F_{t-1} \cup \{R_j(t), (X_i(t), i \in R_j(t))\}, j=1,...,J; t=1,...,T,$$

F_t including and G_{tj} excluding the events which happen in time interval t. F_0 is assumed to represent initial information. If no initial information is available, we can take $F_0 = \emptyset$.

The observation process is given by $\{(R_j(t),(D_i(t),X_i(t),i \in R_j(t)), j=1,...,J), t=1,2,...\}$. Consider then a partially specified statistical model for the observation process. Especially we specify the conditional distributions of $D_i(t)$,

given G_{tj}. It is assumed that the conditional distributions $P(D_i(t)|G_{tj})$ depend on the linear predictor $\beta_{jk}^t{}'X_i(t) + \theta_i$, where β_{jk}^t are parameter vectors common to all of the individuals and θ_i are individual specific parameters. The β_{jk}^t may depend on the origin state, the destination state, and the time interval t.

Several model specifications are possible. A dynamic form of the discrete-time logistic regression model is given by

$$P(D_i(t)|G_{tj}) = \frac{\prod\limits_{k\in M(j)} \exp(\beta_{jk}^t{}'X_i(t)+\theta_i)^{D_{ik}(t)}}{1 + \sum\limits_{k\in M(j)} \exp(\beta_{jk}^t{}'X_i(t) + \theta_i)} Y_{ij}(t) \; . \qquad (2.5)$$

A multivariate probit specification for $P(D_i(t)|G_{tj})$ can also be used instead of (2.5).

Alternative specifications arise if the discreteness of the failure time data is due to the grouping of data from an underlying continuous distribution. One can start with the continuous-time hazard rate or distribution of failure times and then derive discrete-time hazard rates and distributions for grouped data. In general this involves integrals of the density function over the grouping intervals and computations may become laborious but in some cases derivation of the distribution for the grouped model is tractable. Consider for example a proportional hazards model where the transition specific hazard rate for an individual being in state j is given by

$$\lambda_{jk}(t|X_i) = \lambda_{0j}(t) \; \exp(\gamma_{jk}+\beta_{jk}'X_i+\theta_i), \quad k\in M(j)$$

(for uniqueness set one of the γ_{jk}'s equal to zero; the covariates are assumed to be time independent). Then it can be shown that the conditional probabilities $P(D_i(t)|G_{tj})$ are

$$P(D_i(t)|G_{tj}) = (1 - \lambda_{0jt}^{\sum\limits_k c_{ijk}}) \frac{\prod\limits_k c_{ijk}^{D_{ik}(t)}}{\sum\limits_k c_{ijk}} \quad \text{if } \sum\limits_k D_{ik}(t) = 1, \text{ and}$$

$$P(D_i(t)|G_{tj}) = \lambda_{0jt}^{\sum\limits_k c_{ijk}} \quad \text{if } \sum\limits_k D_{ik}(t) = 0 \qquad (2.6)$$

where $c_{ijk} = \exp(\gamma_{jk} + \beta_{jk}'X_i + \theta_i)$. The λ_{ojt} in (2.6) are given by

$$\lambda_{ojt} = \exp(- \int_{a_{t-1}}^{a_t} \lambda_{oj}(u)du)$$

where a_{t-1} and a_t denote lower and upper bound of time interval t $(a_o=0)$. (2.6) may be generalized to include time dependent covariates.

The individual specific parameters are included to account for the effect of unobserved variables (unobserved heterogeneity). A convenient approach is to assume a parametric distribution for the heterogeneity component θ_i (individual parameter) and to estimate the regression coefficients β together with the unknown parameters of the distribution of the heterogeneity component from a 'marginal' likelihood integrating out the unobserved heterogeneity component. Such a model is referred to as a random effect model. If the heterogeneity component is treated as a parameter then the model is referred to as a fixed effect model. The θ_i are incidental parameters (in the sense of Neyman and Scott (1948)) and β_{jk}^t, which is common to all individuals in the sample, is a vector of structural parameters. A basic statistical issue is to develop an estimator for β_{jk}^t that has good properties in this case. A suitable estimation procedure is presented in section 4 for the logistic model.

The use of a fixed effect model will be more appropriate if the individual effects and the included explanatory variables are correlated and if one is not able to give an exact specification of the conditional distribution of θ, given the explanatory variables. Treating the individual effect as an unknown parameter is equivalent to adding a time invariant variable to the set of explanatory variables. Therefore, using a fixed effect model can eliminate the bias arising from the correlation between the unobserved time invariant effects and the included explanatory variables, whereas a random effect inference ignoring the correlation between the effects and explanatory variables can lead to biased estimation. Furthermore, in the fixed effect approach there is no need to postulate a specific distribution of θ. Estimation procedures for the fixed effect approach as well as for the random effect approach are discussed in section 4.

3. Some special cases of interest

3.1 Repeated events of the same kind

This is a one-state process, for example, birth in a fertility history or the lifetimes of an electric appliance until the occurence of a certain defect or break-down. Here, indicator variables $Y_i(t)$ and $D_i(t)$ are defined as follows

$$Y_i(t) = \begin{cases} 1 & \text{if individual is at risk at the beginning of} \\ & \text{time interval } t \\ 0 & \text{otherwise ,} \end{cases}$$

and

$$D_i(t) = \begin{cases} 1 & \text{if individual } i \text{ experiences an event in time} \\ & \text{interval } t \\ 0 & \text{otherwise .} \end{cases}$$

The history of the process up to time t is defined analogously, dropping the subscript j in G_{tj} because there is only one state. The conditional probabilities $P(D_i(t)|G_t)$ are again assumed to be functions of a linear predictor $\beta_t'X_i(t)$ and individual specific parameters θ_i.

A probit specification of the conditional probabilities is

$$P(D_i(t)|G_t) = \Phi[(\beta_t'X_i(t) + \theta_i)(2D_i(t) - 1)] \, Y_i(t) \qquad (3.1)$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution. A logistic regression model is given by

$$P(D_i(t)|G_t) = \frac{\exp(\beta_t'X_i(t) + \theta_i)^{D_i(t)}}{1 + \exp(\beta_t'X_i(t) + \theta_i)} \, Y_i(t) . \qquad (3.2)$$

Other specifications arising from the grouping of a continuous-time model can also be used.

3.2 A two-state model

If there are only two states z_1 and z_2 of interest, e.g. employed - unemployed, then we define the indicator variables $Y_i(t)$ as before and random variables $D_{i1}(t)$ and $D_{i2}(t)$ according to (2.3). The parameter

vector β_2 describes the influence of the covariates or prognostic factors on the conditional transition probabilities $P(D_{i2}(t)|G_t)$ from state z_1 to state z_2, whereas the parameter vector β_1 represents the regression coefficients for the transition from state z_2 to state z_1.

Probit and logit specifications are given by

$$P(D_{i2}(t)|G_t,z_i(t)=z_1) = \Phi[(\beta_{2t}'X_i(t) + \theta_i)(2D_{i2}(t) - 1)] \, Y_i(t),$$

$$P(D_{i1}(t)|G_t,z_i(t)=z_2) = \Phi[(\beta_{1t}'X_i(t) + \theta_i)(2D_{i1}(t) - 1)] \, Y_i(t)$$

$$P(D_{i2}(t)|G_t,z_i(t)=z_1) = \frac{\exp(\beta_{2t}'X_i(t) + \theta_i)^{D_{i2}(t)}}{1 + \exp(\beta_{2t}'X_i(t) + \theta_i)} \, Y_i(t) \, ,$$

$$P(D_{i1}(t)|G_t,z_i(t)=z_2) = \frac{\exp(\beta_{1t}'X_i(t) + \theta_i)^{D_{i1}(t)}}{1 + \exp(\beta_{1t}'X_i(t) + \theta_i)} \, Y_i(t) \, .$$

3.3 Sojourn time in a given state

A special case which is important for practical situations arises if interest is restricted to a certain state and if we investigate the exit rate from this state. Here the end of the first episode is not necessarily the beginning of the second episode and the end of the second episode is usually not the beginning of the third, etc. For example, the successive employment spells of a person can be interrupted by unemployment, further education, illness, etc. In our general model we take this into account by restricting the risk set and the random variables $D_i(t)$ on the state under consideration. $Y_i(t)$ and $D_i(t)$ are defined as follows

$$Y_i(t) = \begin{cases} 1 & \text{if individual i is at risk at the beginning of time} \\ & \text{interval t and } z_i(t)=z \\ 0 & \text{otherwise ,} \end{cases}$$

$$D_i(t) = \begin{cases} 1 & \text{if individual i leaves state z in time interval t} \\ 0 & \text{otherwise .} \end{cases}$$

Specifications of the conditional transition probabilities $P(D_i(t)|G_t)$

are as described in (3.1) and (3.2).

4. Maximum likelihood estimation

The present section deals with the maximum likelihood estimation of the
unknown parameters based on the general model derived in section 2. First
we evaluate a general expression for the likelihood of the observation
process which corresponds to data collected up to time interval T. In
order to keep such a likelihood expression in a manageable form we re-
strict the way in which the law of the process is allowed to depend on
the parameters β and θ. The assumptions generalize those of Arjas (1984).
The resulting likelihood function represents a joint likelihood function
for the structural parameters and the individual specific parameters as
well. One disadvantage is that structural and individual specific para-
meters cannot be estimated consistently from the joint likelihood funct-
ion if the number of time intervals is small. Therefore, our next step
is to derive a conditional likelihood given a suitable sufficient stat-
istic for the individual specific parameter. This conditional likelihood
does not depend on the individual parameters and the structural parameters
can be estimated by maximizing the conditional likelihood function. How-
ever, the conditional approach only applies to the logistic representation
of the conditional transition probabilities. In the last section we invest-
igate the random effect approach which is applicable for all specifications
of the conditional transition probabilities if the distribution of the in-
dividual specific parameters is known.

From the resulting likelihood expression it becomes clear that the estim-
ation procedures are also applicable if some of the first episodes are
left-censored.

4.1 The joint likelihood function

The observation process for the general model of section 2 is

$$\{(R_j(t), (D_i(t), X_i(t), i \epsilon R_j(t)), j=1,..,J), t=1,...,T\} ,$$

corresponding to data collected up to time T, and the likelihood is the
joint probability of the observation process. Using some properties of
conditional probabilities it can be shown that

$$L(\beta,\theta) = \prod_{t \leq T} P((R_j(t), (D_i(t), X_i(t), i\epsilon R_j(t)), j=1,\ldots,J)|F_{t-1}). \quad (4.1)$$

Now we impose the following assumption.

Assumption 1

For each t and (β,θ) the random variables $(R_j(t), (D_i(t), X_i(t), i\epsilon R_j(t))$, $j=1,\ldots,J$, are conditionally independent, given F_{t-1}.

The assumption concerns the conditional independence between the risk sets respectively the individuals who constitute the risk sets. This assumption is likely to hold in practice. Then the likelihood function is given by

$$L(\beta,\theta) = \prod_{t \leq T} \prod_{j=1}^{J} P(R_j(t), (D_i(t), X_i(t), i\epsilon R_j(t))|F_{t-1})$$

$$= \prod_{t \leq T} \prod_{j} P(D_i(t), i\epsilon R_j(t)|R_j(t), (X_i(t), i\epsilon R_j(t)), F_{t-1})$$

$$P(R_j(t), (X_i(t), i\epsilon R_j(t))|F_{t-1}) \quad (4.2)$$

The second term on the right hand side of (4.2) is the joint probability of $R_j(t)$, the individuals at risk in state j during time interval t, and the covariates $\{X_i(t), i\epsilon R_j(t)\}$ measured for these individuals, conditional on the history F_{t-1}. In the following we assume that this probability, given F_{t-1}, does not depend on β and θ.

Assumption 2

For each t, the conditional distribution of $(R_j(t), (X_i(t), i\epsilon R_j(t))$, $j=1,\ldots,J$, given F_{t-1}, does not depend on β and θ.

The assumption states that, given the knowledge contained in F_{t-1}, knowing also the values of $R_j(t)$ and $\{X_i(t), i\epsilon R_j(t)\}$ does not contain additional information about β and θ. Note that in the case where the random variables $R_j(t)$ govern the right censoring of the individuals the assumption implies that such censoring is noninformative (see Kalbfleisch and Prentice (1980), ch. 5.2). For a further discussion of the assumptions see Arjas (1984).

If the assumption does not hold, the likelihood expressions mentioned below can be considered as partial likelihood functions (see Cox (1975)). Otherwise, it becomes necessary to specify the conditional probabilities in the second term on the right hand side of (4.2).

In addition we impose a third assumption which is strictly connected to assumption 1.

Assumption 3

For each risk set $R_j(t)$, $j=1,\ldots,J$, $t=1,\ldots,T$, and for each (β,θ) the random variables $D_i(t)$, $i\epsilon R_j(t)$, are conditionally independent, given G_{tj}.

Assumption 3 again states a conditional independence assumption between the individuals in the sample.

Then the relevant factor of the likelihood which is again denoted by $L(\beta,\theta)$ becomes proportional to the expression

$$L(\beta,\theta) = \prod_{t\leq T} \prod_{j} \prod_{i\epsilon R_j(t)} P(D_i(t)|G_{tj}) \qquad (4.3)$$

The conditional probabilities on the right hand side of (4.3) have still to be specified. For this purpose we can use one of the models discussed in the previous sections.

But if we use (4.3) as a joint likelihood function for the parameters β and θ, a difficulty arises. The parameters β and θ cannot be estimated consistently from this joint likelihood if the number T of time intervals is finite. The reason is that the number of individual specific parameters increases with sample size. Andersen (1973, p. 68-71) considers the binary logit model with T=2 and one structural parameter β. He shows that plim $\hat{\beta}$ = 2β. The same result is obtained for any symmetric distribution, not just the logistic one. Heckman (1981b, p. 187) gives an heuristic argument. He points out that the roots of the likelihood equations involve the joint solution of structural and individual specific parameters. Since estimators of θ_i are necessarily inconsistent, if T is finite, the inconsistency of the estimator for the individual specific parameters is then transmitted to the estimator for the structural parameters.

The inconsistency decreases if T becomes large and in the limit (T→∞) disappears. Therefore, estimation of β (and θ) by maximizing the joint likelihood function (4.3) can be used, if the number of time intervals is moderate or large. But further Monte Carlo studies are needed to determine the size of T such that the maximization of the joint likelihood function performs satisfactory estimates.

4.2 A conditional likelihood function

Our next step is to derive an alternative approach using a conditional likelihood function. The key idea is to base the likelihood function on the conditional distribution of the data, conditioning on a set of sufficient statistics for the individual parameters. But this approach only applies to the logistic model. For this model a sufficient statistic for the individual specific parameter is given by

$$N_i = \sum_{t=1}^{t_i} \sum_k D_{ik}(t) \tag{4.4}$$

where $t_i = \max\{t: Y_{ij}(t)=1 \text{ for some } j\epsilon\{1,\ldots,J\}\}$. N_i is the number of completed spells of individual i.

We consider the conditional probability

$$P((R_j(t), (D_i(t), X_i(t), i\epsilon R_j(t)), j=1,\ldots,J), t=1,\ldots,T \mid N_i, i=1,\ldots,n) \tag{4.5}$$

Since the event counts $D_i(t)$ are part of the event which defines the condition in (4.5), we rewrite (4.5) as the quotient

$$\frac{P((R_j(t), (D_i(t), X_i(t), i\epsilon R_j(t)), j=1,\ldots,J), t=1,\ldots,T)}{P(N_1,\ldots,N_n)} . \tag{4.6}$$

The probability in the nominator of (4.6) is given by (4.3) multiplied by a factor which does not depend on the parameters because of assumptions 1 to 3. Substitution of the logistic representation (2.5) for the conditional probabilities $P(D_i(t)|G_{tj})$ into (4.3) yields

$$L(\beta,\theta) = \prod_{t\leq T} \prod_j \prod_{i=1}^n \frac{\exp(\sum_k \beta_{jk}^t {}'X_i(t) \, D_{ik}(t)Y_{ij}(t) + \theta_i \sum_k D_{ik}(t)Y_{ij}(t))}{(1 + \sum_{k\epsilon M(j)} \exp(\beta_{jk}^t {}'X_i(t) + \theta_i))^{Y_{ij}(t)}}$$

$$= \prod_{i=1}^n \frac{\exp(\sum_t \sum_j \sum_k \beta_{jk}^t {}'X_i(t)Y_{ij}(t) \, D_{ik}(t) + \theta_i \sum_t \sum_j \sum_k D_{ik}(t)Y_{ij}(t))}{\prod_t \prod_j (1 + \sum_{k\epsilon M(j)} \exp(\beta_{jk}^t {}'X_i(t) + \theta_i))^{Y_{ij}(t)}}$$

$$\tag{4.7}$$

The probability in the denominator of (4.6) is given by

$$P(N_1,\ldots,N_n) = \prod_{i=1}^{n} \overbrace{\sum_{\substack{t_i \\ \sum\limits_{t=1}^{} \sum\limits_{k} D_{ik}(t)=N_i}}}^{} \frac{\exp(\sum\limits_{t}\sum\limits_{j}\sum\limits_{k}\beta_{jk}^t{}'X_i(t)Y_{ij}(t)D_{ik}(t) + \theta_i N_i)}{\prod\limits_{t}\prod\limits_{j}(1 + \sum\limits_{k\epsilon M(j)}\exp(\beta_{jk}^t{}'X_i(t) + \theta_i))^{Y_{ij}(t)}} \tag{4.8}$$

and the conditional likelihood function which is denoted by $CL(\beta)$ is obtained by dividing (4.7) by (4.8)

$$CL(\beta) = \prod_{i=1}^{n} \frac{\exp(\sum\limits_{t}\sum\limits_{j}\sum\limits_{k\epsilon M(j)}\beta_{jk}^t{}'X_i(t)Y_{ij}(t)D_{ik}(t))}{\sum\limits_{\sum\limits_{t}\sum\limits_{k}D_{ik}(t)=N_i}\exp(\sum\limits_{t}\sum\limits_{j}\sum\limits_{k\epsilon M(j)}\beta_{jk}^t{}'X_i(t)Y_{ij}(t)D_{ik}(t))} \tag{4.9}$$

The conditional likelihood function (4.9) only depends on the structural parameters β, and does not depend upon the individual specific parameters. Hence standard asymptotic theory applies. The conditional ML-estimator of β is consistent and asymptotically normally distributed provided that the individual parameters and the conditional likelihood function satisfy regularity conditions (see Andersen 1973).

Note that individuals with $N_i=0$ or $N_i=t_i$ where $t_i=\max\{t: Y_{ij}(t)=1$ for some $j\epsilon\{1,\ldots,J\}\}$ do not contribute any information to the conditional likelihood (4.9), since for these values of N_i nominator and denominator on the right hand side of equation (4.9) are equal. Therefore, the number of individuals who have one or more completed spells should be at least moderate.

We must keep in mind that the conditional likelihood method is only helpful in a logit model. It is not generally possible to find minimum sufficient statistics for the individual specific parameters which are independent of the structural parameters and which have a smaller dimension than the sample size. This is possible if the distribution is a member of the exponential family like the logistic parametrization of the multinomial distribution. Therefore, conditional likelihood methods are not a general approach in fixed effect models, but if the logistic represent-

ation is appropriate, the conditional approach has some advantages. It does not require a specification for the distribution of θ. If one makes such an assumption, the distribution of the heterogeneity θ conditional on the observed covariates X is needed, and in general this should be allowed to depend upon the observed covariates. If there was omitted variable bias before introducing θ, and if one mistakenly models θ as independent of X, then the resulting estimator based on the 'marginal' likelihood (see next section) will also be biased. The fixed effect approach presented here has the advantage of allowing for a very general relationship between θ and X.

Note that the appropriateness of the logistic representation can in principle be tested by one of the specification tests for the multinomial logit model described by Hausman and McFadden (1984) in a choice theoretic context.

4.3 A marginal likelihood function

In this section we discuss an alternative approach assuming that the individual specific parameters follow a distribution. The individual specific parameter θ, the heterogeneity, is not observable. Let $G(\theta)$ denote the (marginal) distribution of θ. In this case the probabilities $P(D_i(t)|G_{tj},\theta_i)$ are conditional probabilities, given G_{tj} and θ_i, and the resulting likelihood function (4.3) is also conditional on the individual specific components θ_i. Introducing the indicator variables $Y_{ij}(t)$ as defined in (2.1) we can rewrite the contribution of individual i to the likelihood expression in (4.3)

$$L_i(\beta|\theta_i) = \prod_{t=1}^{T} \prod_{j} P(D_i(t)|G_{tj},\theta_i)^{Y_{ij}(t)} . \tag{4.10}$$

If it is possible to specify $G(\theta)$ as a member of a parametric class of probability distributions, estimation of the structural parameters can be based on the marginal distribution of the observation process integrating out the individual component θ. The marginal likelihood function is denoted by $ML(\beta,\gamma)$ where γ is the parameter vector determining the distribution of θ. It is given by

$$ML(\beta,\gamma) = \prod_{i=1}^{n} \int \prod_{t=1}^{T} \prod_{j} P(D_i(t)|G_{tj},\theta)^{Y_{ij}(t)} dG(\theta) . \tag{4.11}$$

The marginal likelihood function is a function of β and the unknown parameters γ of the population distribution $G(\theta)$. Maximization of this likelihood function will, under weak regularity conditions, give consistent and asymptotically normally distributed estimators for β and γ.

Note that in this approach the population distribution $G(\theta)$ is assumed to be known except for a finite number of parameters. Furthermore, if θ_i and $X_i(t)$ are correlated, we have to specify the joint distribution of (θ_i, X_i') in order to obtain consistent estimates of structural parameters. A convenient possibility in analogy to the linear model case is to assume that the dependence is only via a linear regression function (Chamberlain (1980, 1984))

$$\theta_i = \pi'X_i + \varepsilon_i \quad , \quad i=1,\ldots,n \quad (4.12)$$

where $X_i'=(X_i(1)',\ldots,X(t_i)')$, t_i as defined in (4.4), and where ε_i is independent of X_i. We assume that the ε_i are independent and identically distributed with distribution function $H(\varepsilon)$. Substitution of (4.12) into (4.10) and (4.11) yields a marginal likelihood function which is appropriate if the heterogeneity component is correlated with the observed covariates. This seems to be rather the rule than the exception.

For illustration let us consider the special case described in section 3.1 with the probit specification (3.1). Then, allowing for correlation between θ_i and X_i and using (4.12) the marginal likelihood function is given by (the parameter vector determining $H(\varepsilon)$ is denoted by α)

$$ML(\beta,\alpha) = \prod_{i=1}^{n} \int \prod_{t=1}^{T} \prod_{j} \{\Phi((\beta_t'X_i(t)+\pi'X_i+\varepsilon)(2D_i(t)-1))\}^{Y_{ij}(t)} dH(\varepsilon) \quad (4.13)$$

Note that sometimes identification problems may arise especially if the parameters β are assumed to be time independent.

Finally we mention alternative approaches of Liang and Zeger (1986) and Stiratelli, Laird and Ware (1984).

Liang and Zeger propose methods for longitudinal data (not for event history or failure time data) only assuming a functional form for the marginal distribution at each time corresponding to $P(D_i(t))$ in the present paper. The marginal distribution is assumed to belong to the family of generalized linear models. In addition, a covariance structure for $(D_i(1),\ldots,D_i(T))$ is assumed but this covariance structure across

time is treated as a nuisance. Then they derive estimating equations
similar to the quasi-likelihood approach (see, for example, McCullagh
(1983)) and investigate asymptotic properties of the estimators of the
regression coefficients.

Stiratelli, Laird and Ware (1984) consider the special case of longitu-
dinal data with binary outcomes. They split up the set of covariates into
two sets. The first set contains the covariates which vary over time
and in the second set are the covariates which are fixed. The fixed co-
variates are denoted by x_i and the time varying covariates are denoted
by z_{it}. Furthermore, let λ_i denote the T-vector of logits for individual i.
Then, Stiratelli, Laird and Ware (1984) investigate a two-stage approach
where at stage 1 they let

$$\lambda_i = X_i \beta + Z_i \alpha_i$$

with suitable defined matrices X_i and Z_i, and at stage 2 they assume that
α_i is multivariate normal with expectation 0 and covariance matrix Σ.
These assumptions define a general mixed model for the logits of the
response probabilities and one could try to carry over this model into
the event history or failure time context using the EM algorithm or
empirical Bayes strategies for estimation.

References

Andersen EB (1973) Conditional inference and models for measuring.
Kopenhagen

Andersen PK, Gill RD (1982) Cox's regression model for counting processes:
a large sample study. Annals of Statistics 10:1100-1120

Allison P (1982) Discrete-time methods for the analysis of event histories.
In: Leinhardt S (ed) Sociological Methodology, pp. 61-98

Arjas E (1984) Logistic regression in modelling hazard: a fully parametric
approach. Preprint, University of Oulu, Finland

Arjas E, Haara P (1986) A logistic regression model for hazards: asympt-
otic results. Scandinavian Journal of Statistics 13 (forthcoming)

Blossfeld HP, Hamerle A, Mayer KU (1986) Ereignisanalyse: Statistische
Theorie und Anwendungen in den Wirtschafts- und Sozialwissenschaften.
Campus, Frankfurt

Borgan Ø (1984) Maximum likelihood estimation in parametric counting
process models, with applications to censored failure time data. Scand-
inavian Journal of Statistics 11:1-16, Correction 11:275

Breslow NE (1974) Covariance analysis of censored survival data. Biometrics
30:89-99

Chamberlain G (1980) Analysis of covariance with qualitative data. Review of Economic Studies 17:225-238

Chamberlain G (1984) Panel data. Handbook of Econometrics, Vol. II (Z. Griliches & M Intriligator, Eds.), North-Holland, Amsterdam

Chamberlain G (1985) Heterogeneity, omitted variable bias, and duration dependence. In: Heckman J, Singer B (eds) Longitudinal analysis of labor market data. Cambridge University Press, Cambridge

Cox DR (1975) Partial likelihood. Biometrika 62:269-276

Cox DR, Oakes D (1984) Analysis of survival data. Chapman and Hall, London

Haberman SJ (1977) Maximum likelihood estimates in exponential response models. Annals of Statistics 5:815-841

Hamerle A (1984) Zur statistischen Analyse von Zeitverläufen. Diskussionsbeitrag Nr 180, Universität Regensburg

Hamerle A (1985) Regressionsmodelle für gruppierte Verweildauern und Lebenszeiten. Zeitschrift Operations Research 29, B:243-260

Hanefeld U (1984) Das Sozio-ökonomische Panel - eine Längsschnittstudie für die Bundesrepublik Deutschland. Vierteljahreshefte zur Wirtschaftsforschung 4/84:391-406

Hausman J, McFadden D (1984) Specification tests for the multinomial logit model. Econometrica 52:1219-1240

Heckman J (1981a) Statistical models for discrete panel data. In: Manski CF, McFadden D (eds) Structural analysis of discrete data with econometric applications, MIT Press, Cambridge, Mass.:114-178

Heckman J (1981b) The incidental parameter problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In: Manski CF, McFadden D (eds) Structural analysis of discrete data with econometric applications. MIT Press, Cambridge, Mass.:179-195

Heckman J, Singer B (1984) Econometric duration analysis. Journal of Econometrics 24:63-132

Hujer R, Schneider H (1986) Ökonometrische Ansätze zur Analyse von Paneldaten: Schätzung und Vergleich von Übergangsratenmodellen. Manuskript, Frankfurt

Kalbfleisch JD, Prentice RL (1980) The statistical analysis of failure time data. Wiley, New York

Laird N, Olivier D (1981) Covariance analysis of censored survival data using log-linear analysis techniques. Journal of the American Statistical Association 76:231-240

Lawless JF (1982) Statistical models and methods for lifetime data. Wiley, New York

Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika 73:13-22

Mantel N, Hankey BF (1978) A logistic regression analysis of response-time data where the hazard function is time dependent. Communications in Statistics A7:333-347

McCullagh P (1983) Quasi-likelihood functions. Annals of Statistics 11:59-67

Neyman J, Scott EL (1948) Consistent estimates based on partially consistent observations. Econometrica 16:1-32

Prentice RL, Gloeckler LA (1978) Regression analysis of grouped survival data with applications to breast cancer data. Biometrics 34:57-67

Stiratelli R, Laird N, Ware J (1984) Random effect models for serial observations with binary responses. Biometrics 40:961-971

Thompson WA Jr (1977) On the treatment of grouped observations in life studies. Biometrics 33:463-470

Tuma NB, Hannan MT (1984) Social dynamics: Models and methods. Academic Press, New York

Unobserved heterogeneity in models of unemployment duration

Heinz P. Galler and Ulrich Poetter

1. Introduction

In analyses of unemployment duration data, it is of substantial interest to discriminate between true negative duration dependence of the exit rate from unemployment, caused for instance by dequalification, and spurious negative duration dependence due to unobserved heterogeneity of the observations. Beside the different implications for labor market policy, it is also important to separate the impact of unobserved heterogeneity from true duration dependence for statistical reasons. As demonstrated by Heckman and Singer (1984), parameter estimates may be rather sensitive with regard to the assumptions introduced for the distribution of unobserved heterogeneity in the population. At the same

time, findings by Trussell and Richards (1985) indicate that estimates also depend on the specification of duration dependence adopted. This makes the model specification a crucial point in empirical analyses of duration data.

In the case of unemployment duration it is rather difficult to derive the structural form of the hazard rate from (economic) theory. The search model commonly used by economists leads to rather complex nonlinear relations that cannot be solved analytically without further simplifications. Approximations have to be applied to obtain an operable specification suited for empirical work. However, since different approximations can be applied, the question is raised how to deal appropriately with true duration dependence and unobserved heterogeneity in such simplified models.

2. Search models of unemployment duration

The basic search model of unemployment duration has been developed by Mortensen (1970) and McCall (1970). In recent years, the approach of Kiefer and Neumann (1979a,1979b,1981) has been used by several researchers in empirical studies. It is based on the analysis of a rational individual's decision to accept or to reject job offers that are assumed to become available at random with a known probability distribution.

In the standard search model, job offers are assumed to become available to a given individual at a constant rate. Jobs are described by the wage offered which is modelled as a random variable with a known conditional probability distribution. Under fairly general conditions, an optimal decision rule for an unemployed individual i at time t then consists in fixing a reservation wage $V(i,t)$ and to accept

any offer with a wage w exceeding this reservation level
(cf. Hall et.al. 1976).

In general, the reservation wage is defined by the net bene-
fits expected from further search: a rational individual
will accept a job offered, if its expected benefits exceed
those of further search. If this is true, the conditional
probability that an individual accepts a job offered and
leaves unemployment equals the probability that a wage offer
exceeding the reservation wage is received. The exit rate
from unemployment then is defined as the product of the rate
at which feasible offers become available and the conditio-
nal probability that the job will be accepted. If $F(w|i,t)$
is the probability distribution of job offers faced by in-
dividual i at time t and $g(i,t)$ is the rate at which job
offers become available, the exit rate $h(i,t)$ from unem-
ployment in general is defined as a function of the reser-
vation wage $V(i,t)$ by:

$$(1) \quad h(i,t) = g(i,t) [1 - F(V(i,t)|i,t)]$$

For the basic search model, usually stationarity is assumed
as well as an infinite time horizon. In this case, the re-
servation wage remains constant over time. However, this
specification is rather restrictive. Even if short periods
are considered for which homogeneity of the environment may
be assumed, reservation wages in general will not be con-
stant. A finite time horizon for job search will result in
varying reservation wages. Beside that, the assumption of
stationarity of the job offer distribution appears to be
questionable. If work contracts are regarded as the result
of search activities by both employers and employee (eg.
Galler 1985), the rate at which job offers become available
to an individual may depend on the duration of unemployment.
Rational employers searching for a worker to fill a given
vacancy will define some minimum requirements with regard to

qualification and other characteristics. If the duration of unemployment is used as a screening device, the chances of an unemployed individual to be offered a job probably will decline with increasing duration of unemployment even if the total number of jobs available remains constant.

Even for the simplified approach based on the assumption of stationarity, in general no direct solution can be obtained for the reservation wage $V(i,t)$. In a discrete time approach Kiefer and Neumann (1981, p.174) start from a general representation of the reservation wage as a function of the variables determining the wage offer distribution, the net transfer received, the discount rate and the duration of unemployment. Then, a linear approximation is applied to obtain an operable representation of the reservation wage relation that can be estimated from the empirical data by probit or logit procedures.

A similar approach can be used for a nonstationary continuous time formulation of the model. Since for time continuous models the integrated hazard rate must be derived, it is preferable to start from a general representation of the logarithm of the hazard rate and then to apply a linear expansion. In general, the exit rate from unemployment will depend on the rate $g(i,t)$ of offers becoming available, on the conditional wage offer distribution $F(w|i,t)$, on the reservation wage $V(i,t)$ and on current transfer income $z(i,t)$. Additionally, expectations with regard to future job offers and transfers will also determine the reservation wage, if no stationarity is assumed. If the impact of expectations is collapsed into a variable $Q(i,t)$, a general representation of the hazard rate is:

$$(2) \quad \ln h(i,t) = h^*\{ g(i,t), F(w|i,t), z(i,t), Q(i,t) \}$$

Since expectations are formed on the basis of information
available, the terms entering the hazard rate (2) may be
modelled as functions of the observed characteristics x(i,t)
of the individual, of labor market indicators y(t), of the
duration of unemployment t, as well as unobserved variables
e(i,t). By a linear expansion with regard to the observed
variables x(i,t) and y(t) and some function k(t) represent-
ing duration dependence of unemployment, the following rela-
tion is obtained with an error term R(i,t), which describes
the approximation error and the impact of the unobserved
variables:

$$(3) \quad \ln h(i,t) = h_o + \alpha' x(i,t) + \beta' y(t) + \pi k(t) + R(i,t)$$

Parameter estimates based on such a linear approximation
provide a basis for inferences about the structure of the
true relation. However, without additional assumptions with
regard to the error R(i,t) the model cannot be estimated.

3. Specifications of unobserved heterogeneity

By definition, the error term of the simplified model repre-
sents deviations of the model from the true relation. It may
be interpreted as a specification error caused by the omis-
sion of variables relevant for explanation but not included
into the model. To some extend misspecification will occur
even if all relevant variables are included into the model
but an approximation to the true functional form of the re-
lation is used. Such specification errors imply deviations
of the true hazard rate for the individual from the estimate
derived from the model.

In the most simple formulation of the duration model, the
heterogeneity term R(i,t) is neglected. In this case omitted

covariates, departure from proportionality or an inadequate
model of duration dependence may lead to a severe bias in
the estimates of the form of duration dependence and the in-
fluence of covariates. When the "true" model is a mixture of
Weibull models but a simple Weibull model is used for esti-
mation, the estimated form parameter as well as the esti-
mated parameters of the covariates will be downward biased
(Lancaster 1985).

A less restrictive specification is obtained, if the form of
the baseline hazard is not explicitly modelled and inference
is drawn from Cox' partial likelihood. However, the Cox
model does not provide estimates of the form of duration de-
pendence. For this purpose, alternative approaches like that
by Breslow have to be used (cf. Kalbfleisch/Prentice 1980,
p.78).

In the presence of unobserved heterogeneity the estimates
from a partial likelihood approach are downward biased. The
magnitude of the bias is determined by the influence of the
omitted variables on the duration in a state (Struthers/
Kalbfleisch 1986). Departure from proportionality results in
a bias, whose direction depends on the form of the "true"
underlying model, but the relative importance of the covari-
ates remains unchanged, at least to a first order approxima-
tion (Solomon 1984).

One way to treat the error term more explicitly is to model
it by a time invariant constant that is specific to the in-
dividual observation. If these terms are supposed to be re-
alizations of a random variable, this reduces to the ap-
proach to unobserved heterogeneity usually adopted. It im-
plies the assumption that the individual hazard rates are
proportional up to an unknown constant heterogeneity factor
given the values of the explanatory variables. The marginal
distribution of such models will in general no longer belong
to the class of proportional hazard models. The deviations

of the empirical duration distribution from the form of time
dependence assumed in the model can be used for inference
about the distribution of unobserved heterogeneity in the
population.Thus a constant heterogeneity term can be used to
model individual heterogeneity or departure from the sup-
posed form of time dependence.

If a parametric form of the probability distribution of the
heterogeneity term is assumed, such a model can be estimated
directly by maximum likelihood methods. Moreover the combi-
nation of a parametric hazard model and a parametric hetero-
geneity specification yields a flexible class of duration
distributions and allows for the representations of various
aspects of heterogeneity. If sufficient prior information on
the properties of heterogeneity is available, the methods of
Hougaard (1984, 1986) can be used to select an appropriate
distribution for the heterogeneity term. However, almost no
restrictions can be deduced from theory with regard to the
distribution of unobserved heterogeneity.

Considering the nonlinear form of the underlying structural
relations, a parametric assumption like the log-normal or
the gamma distribution may not be appropriate. Even if such
an assumption is valid for the distribution of the unob-
served characteristics in the population, the nonlinear
structure of the true relation may result in a nonstandard
distribution of the error terms of the approximate relation.
As a consequence, parametric estimators based on standard
distributional assumptions may yield biased estimates as
demonstrated by Heckman and Singer (1984). This makes a non-
parametric approach to unobserved heterogeneity attractive.

Semi-parametric models with a parametric specification of
the baseline hazard and a nonparametric specification of the
heterogeneity term pose problems of identifiability and re-
quire nonstandard maximum likelihood estimation techniques.
If a large number of nuisance parameters is used to repre-

sent individual heterogeneity, maximum likelihood estimators need not be consistent (Neyman/Scott 1948). But for specific types of models of the hazard rate using some additional assumptions on the distribution of heterogeneity and the admissible parameters, Heckman/Singer (1984) were able to proof the identifiability of such a model and the consistency of a semiparametric maximum likelihood estimator. Simulation results suggest good properties in the estimation of structural parameters, while estimates of the distribution of heterogeneity even in large samples do not recover the underlying "true" distribution very well (Heckman/ Singer 1984).

However, in the nonparametric Heckman-Singer approach to unobserved heterogeneity, some basic problems remain. In their empirical work Trussell and Richards (1985) report the strong dependence of the semiparametric estimators on the choice of a parametric baseline hazard. One way to circumvent this difficulty is to adopt a nonparametric specification of both duration dependence and of the heterogeneity term. In the case of a heterogeneity term with finite expectation and in the presence of at least one regressor, Elbers and Ridder (1982) demonstrated the identifiability of such a model.

An inspection of their proof suggests that a large amount of (exact) data would be needed to recover the form of the hazard, the influence of the covariates and the functional form of the heterogeneity term simultaneously. An empirical implementation of their proof strategy would require numerical differentiation and integration. These procedures are very sensitive to small perturbations in the values of the estimated survival function. Furthermore the knowledge of the complete survival function for all values of the covariates is needed. Given the limitations of the data, it seems not to be possible in most cases to estimate both duration dependence and heterogeneity nonparametrically.

A flexible strategy may be to adopt a partially parametric dummy variable approach as an approximation to a nonparametric formulation of duration dependence (eg. Trussell and Richards 1985). In such a model duration dependence is modelled by dummy variables that represent given predetermined time intervals of the duration of unemployment. This is not a fully nonparametric approach since the intervals have to be fixed a-priori. But it allows a rather flexible specification of duration dependence so that misspecifications can be avoided to a large extend. This is especially true if many dummy variables are used.

Since a dummy variable specification implies some restrictions for the form of duration dependence, it can in principle be combined with a nonparametric specification of unobserved heterogeneity. Thus it provides a flexible specification for both duration dependence and unobserved heterogeneity. However, one would expect unobserved heterogeneity to be empirically of less importance for such a partially parametric model than for other, more specific specifications of duration dependence since by dummy variables a better fit to the empirical distribution of unemployment durations can be achieved. Especially if a comparatively large number of dummy variables is employed identification of unobserved heterogeneity, while theoretically possible, would require a rather large amount of exact data.

4. The data base and models used for estimation

As an example for the consequences of different specifications of duration dependence and unobserved heterogeneity, different version of a hazard rate model for unemployment durations have been estimated. As a data base, the information of the first two waves of the German Socio-Economic Panel Study has been used. This panel study has been started

in 1984 by the Sonderforschungsbereich 3 to collect longitu-
dinal data on the population living in private households in
West Germany (cf. Hanefeld 1984). Presently, experience with
these data is still limited. Only a few analyses have been
conducted yet. Therefore the analyses reported here are of
exploratory character.

Beside other information, in each wave of the Socio-Economic
Panel Study the unemployment status in the previous year is
recorded for each person in the sample on a monthly basis by
retrospective interviewing. Joining the data from the first
two waves for each individual, a period of 24 month is cov-
ered starting from January 1983. Within this period, the
start and the end of individual spells of work, unemploy-
ment, schooling or non-participation in the labor force can
be dated up to the month. As explanatory variables, rich in-
formation on individual characteristics is available as well
as on the family and the household. However, because of pri-
vacy issues, presently no regional information is supplied
that would allow to use regional labor market indicators as
explanatory variables.

Since female labor market behavior depends on rather in-
volved decisions with regard to labor force participation
and household activities, the analysis was restricted to un-
employed men. To avoid problems of left censoring, only
those individuals were include into the sample, that had be-
come unemployed during the period considered.

In the data base, the start and the end of an unemployment
spell is defined by comparing the individual's status in two
consecutive months. Because only the month is recorded, the
start and the end of each spell has been dated to the middle
of the respective month except for those individuals who re-
mained in unemployment only for one month. For these cases a
duration of a quarter of a month was assumed. Observations
were marked as right censored after the month of November,

if they remained in the same status in December but no
information was available on the status during January of
the following year. For the same reason, spells starting in
December but not followed in January were excluded from the
sample since it is not possible to decide whether the spell
had been completed during December or whether the observa-
tion is right censored.

For the present analysis three model specifications have
been used that differ only in the specification of duration
dependence of the exit rate from unemployment. First, an
exponential model with no duration dependence has been con-
sidered. In the second model duration dependence of the
Weibull type has been assumed. The third specification uses
dummy variables μ that indicate durations of 1-3 months and
more than 3 months respectively:

(4) Model (1): $\ln[h(i,t)] = \alpha'x(i,t)$

Model (2): $\ln[h(i,t)] = \alpha'x(i,t) + (\Phi-1)\ln(t) + \ln(\Phi)$

Model (3): $\ln[h(i,t)] = \alpha'x(i,t) + \mu(t)$

Only a few explanatory variables have been included into the
model. First a dummy variable is used to indicate non-German
nationality. Age is represented by a set of dummy variables
for different age brackets. For education, dummy variables
are introduced for individuals without formal professional
training and or with university degrees respectively. In the
present specification, unemployment transfers are only re-
presented by a dummy variable that takes the value one for
those months in which unemployment compensation was re-
ceived. No variables related to the family size or to the
marginal utility of income have been included. The same is
true for labor market indicators, since no regional informa-
tion was supplied and indicators on the national level

showed too little variation to be used as explanatory variables.

Additionally, a dummy variable was introduced for the month of December 1983 to represent the effect of recall errors. Since the information on the two years has been collected in different waves of the panel study, obviously some inconsistencies occur. A significant proportion of respondents who had reported in the first wave to have been unemployed in December 1983 did not report unemployment for January 1984 in the second wave. To some extend this probably is caused by recall errors. The dummy variable is intended to catch this effect. For the same reason, observations starting in January 1984 have been excluded from the sample since at least some portion of them is left censored due to under-reporting in the first wave.

To assess the sensitivity of the estimates with regard to unobserved heterogeneity, all three models have been estimated both under the assumption of no unobserved heterogeneity and for a nonparametric specification of unobserved heterogeneity based on the Heckman-Singer approach. For unobserved heterogeneity, estimation started with a sufficiently large number of support points $\Theta(i)$ of the mixing distribution. Then the number of points was reduced, if either the probability mass of a point approached zero or two points of support converged during the iteration process.

5. Estimation results

The model specifications without unobserved heterogeneity have been estimated using standard maximum likelihood procedures. For numerical maximization of the log-likelihood function the quadratic-hill-climbing procedure as developed by Goldfeld and Quandt (1972) was used. For nonparametric

estimation of the mixing distribution for unobserved hetero-
geneity, basically the EM-algorithm has been applied (cf.
Redner/Walker 1984). This approach proved to require rather
extensive computations.

For the models without unobserved heterogeneity, the parame-
ter estimates are reported together with the t-ratios. How-
ever, for the semiparametric models no computational feasi-
ble lower bound for the variance of the estimators is avail-
able. Note that the reported t-ratios have no strict justi-
fication either when computed under the wrong model.

The parameter estimates for the specification without unob-
served heterogeneity are similar to the findings of other
studies. The exit rate from unemployment is lower for
foreigners than for German nationals. Younger individuals
leave unemployment faster than the reference group in the
middle age bracket while older persons show a significantly
lower exit rate. Men without professional qualifications
leave unemployment slower than the reference group with for-
mal professional training while the group with university
degrees shows substantially higher exit rates. But from a
theoretical point of view, one would rather expect more
qualified individuals to search longer for a job. The exit
rate is substantially lower if unemployment benefits are re-
ceived. Finally, the dummy variable for December 1983 shows
a significant positive effect of recall errors on the exit
rate.

The parameter estimates for the covariates prove to be
rather robust with regard to the specification of duration
dependence assumed. Except for the constant, there are only
slight differences in the estimates between the three model
specifications. This result is in accordance with the find-
ings of Trussell and Richards (1985).

Table 1: Parameter estimates for hazard rate models of unemployment
duration for unemployed men under 65 years in West Germany
(Sample size n=441, t-values in parentheses)

Model	No unobs. Heterogeneity			With unobs. Heterogeneity		
	(1)	(2)	(3)	(1)	(2)	(3)
Log-Likelihood	-71.9	-65.7	-44.9	-62.3	-51.9	-44.9
Foreigner	-.1317	-.1163	-.1036	-.1086	-.1440	-.1035
	(.93)	(.82)	(.73)	-	-	-
Age under 25	.6819	.6463	.6105	.8401	1.3314	.6108
	(3.77)	(3.57)	(3.37)	-	-	-
25 - 34	.3614	.3553	.3324	.4613	.2343	.3326
	(1.84)	(1.81)	(1.69)	-	-	-
45 - 54	-.1280	-.0766	-.0839	-.0590	-.2995	-.0837
	(.52)	(.31)	(.34)	-	-	-
55 - 64	-.9806	-.8850	-.8307	-1.1312	-2.2432	-.8309
	(2.62)	(2.37)	(2.22)	-	-	-
Education						
no prof. tr.	-.3697	-.3760	-.3844	-.5183	-1.7249	-.3846
	(2.46)	(2.50)	(2.56)	-	-	-
university	.6804	.5779	.4970	.5942	-.1571	.4969
	(2.84)	(2.41)	(2.07)	-	-	-
Unempl. benef.	-.1989	-.2071	-.2102	-.1992	.2432	-.2103
	(1.51)	(1.57)	(1.56)	-	-	-
December '83	.9442	.9096	.8868	.9244	.7532	.8868
	(5.15)	(4.94)	(4.82)	-	-	-
Constant (Mean)	.4051	.3247	1.0607	.4616	1.7649	1.0604
	(2.40)	(1.91)	(5.77)	-	-	-
Weibull ($\$$)	-	.8483	-	-	1.9167	-
	-	(3.39)	-	-	-	-
Duration						
μ_1: 1-3 m.	-	-	-.8942	-	-	-.8939
	-	-	(5.76)	-	-	-
μ_2: 4+ m.	-	-	-.9584	-	-	-.9576
	-	-	(6.66)	-	-	-
Mixing Distrib.						
1. Point θ_1	-	-	-	-.8862	-2.8868	-.0245
p_1	-	-	-	.2272	.2989	.5499
2. Point θ_2	-	-	-	-.1794	-.2583	.0299
p_2	-	-	-	.5713	.4151	.4501
3. Point θ_3	-	-	-	1.5077	3.3919	-
p_3	-	-	-	.2015	.2860	-

Source: Socio-Economic Panel Study, first and second wave

Both the Weibull specification (2) and the dummy variable representation (3) show significant negative duration dependence of the exit rate if no unobserved heterogeneity is introduced. This is consistent with the results of other empirical studies (eq. Egle 1979). However, at least from conventional economic search theory, one would rather expect the exit rate to rise with prolonged unemployment since unemployment benefits decrease as well as expected income and individual wealth. But eventually other relationships might be derived if screening by potential employers dependent on unemployment duration is considered.

If unobserved heterogeneity is introduced, quite different effects occur. For the dummy variable specification of duration dependence (3) almost no unobserved heterogeneity is found. In the estimation process, the mixing distribution degenerates to one point equal to the constant term of the corresponding model without unobserved heterogeneity. The parameter estimates are practically identical to those for the simple model. The remaining differences result from the slow convergence properties of the EM-algorithm.

Apparently, the step function assumed for duration dependence fits the data sufficiently well so that no variation remains that could be attributed to unobserved heterogeneity. However, the partially parametric specification of duration dependence is not very well suited either for a combination of positive duration dependence with a mixing distribution. Since for each duration interval a mixture of constant hazard rates is assumed, the mixture will imply negative duration dependence within each interval. Thus a mixture of increasing step functions results in a rather peculiar form of the duration dependence of the hazard rate. This may also be the reason for the estimates to show negative duration dependence.

In contrast, for the specification without any duration
dependence substantial variation of the heterogeneity term
is estimated. This is plausible since introducing unobserved
heterogeneity allows a better fit to the observed distribu-
tion of durations than it can be achieved by a simple expo-
nential model with a constant rate. In comparison to an ex-
ponential model, the density function of a mixture of expo-
nentials is steeper for short durations and flatter for
longer ones. The largest changes in the parameter estimates
occur for the age specific dummy variables. The age specific
differentials of the hazard rate increase as the coeffi-
cients estimated for the lower age brackets rise while those
for the older individuals are reduced further. A similar ef-
fect can be observed for the impact of education.

As compared to the other two models, the changes in the
parameter estimates are largest for the Weibull specifica-
tion. The age specific differentials increase substantially.
However, the most important effects occur for the impact of
education, of unemployment benefits and for the Weibull pa-
rameter itself. While for the specification without unob-
served heterogeneity significant negative duration depen-
dence has been estimated ($\Phi<1$), controlling for unobserved
heterogeneity results in strong positive duration dependence
($\Phi>1$). This resembles the results of Heckman and Singer
(1984) based on the Kiefer-Neumann data. Since positive du-
ration dependence may appear more plausible from a theoreti-
cal point of view, this might be regarded as a result in
favour of a nonparametric specification of unobserved het-
erogeneity.

The same is true for the negative coefficient estimated for
high qualification as compared to the positive value derived
for the specification without unobserved heterogeneity.
Lower exit rates for highly qualified individuals are in ac-
cordance with economic theory. In contrast, the positive
estimate for the impact of unemployment benefits does not

agree with prior theoretical beliefs. Here, the negative
value of the other models appears to be more plausible. How-
ever, as the t-ratios for the model without unobserved het-
erogeneity suggest, this result may not be significant.

In general, the estimates appear to support the argument by
Heckman and Singer (1984) that a nonparametric approach to
unobserved heterogeneity will result in less biased parame-
ter estimates. However, such conclusions should be drawn
with precaution. The rather large values of the coeffi-
cients, the sign-change of two coefficients and the value of
the Weibull parameter give rise to the question whether the
estimates represent true causal dependencies or eventually
reflect some peculiarities of the model or of the data.

A Weibull model with positive duration dependence implies a
unimodal probability density of durations with the modal
value depending on the Weibull parameter given the values of
the other explanatory variables. If for instance due to the
process of data collection, the observed durations concen-
trate around some specific values, a mixture of Weibull mod-
els with positive duration dependence may fit the empirical
frequency distribution better than a model with negative du-
ration dependence. Thus positive duration dependence eventu-
ally may reflect deficiencies of the data rather than a true
causal relationship.

To get an impression of the distribution of durations im-
plied by the estimates, the probability densities have been
computed from the model conditional on the different esti-
mates for the heterogeneity parameter using the sample val-
ues of the explanatory variables. These conditional proba-
bility densities then have been aggregated to the marginal
density for the sample that is compared to a nonparametric
life table estimate of the duration distribution.

Duration densities for mixture classes

Weibull modell with 3 support points
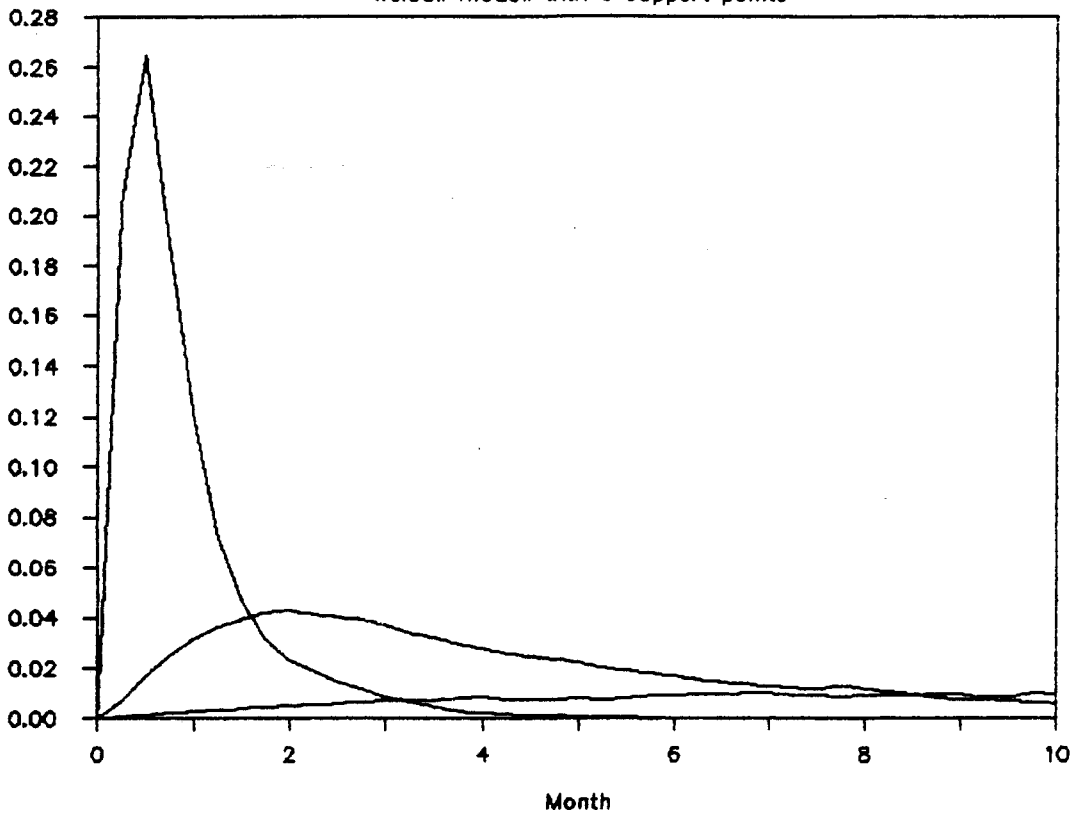
Duration densities conditional on support points
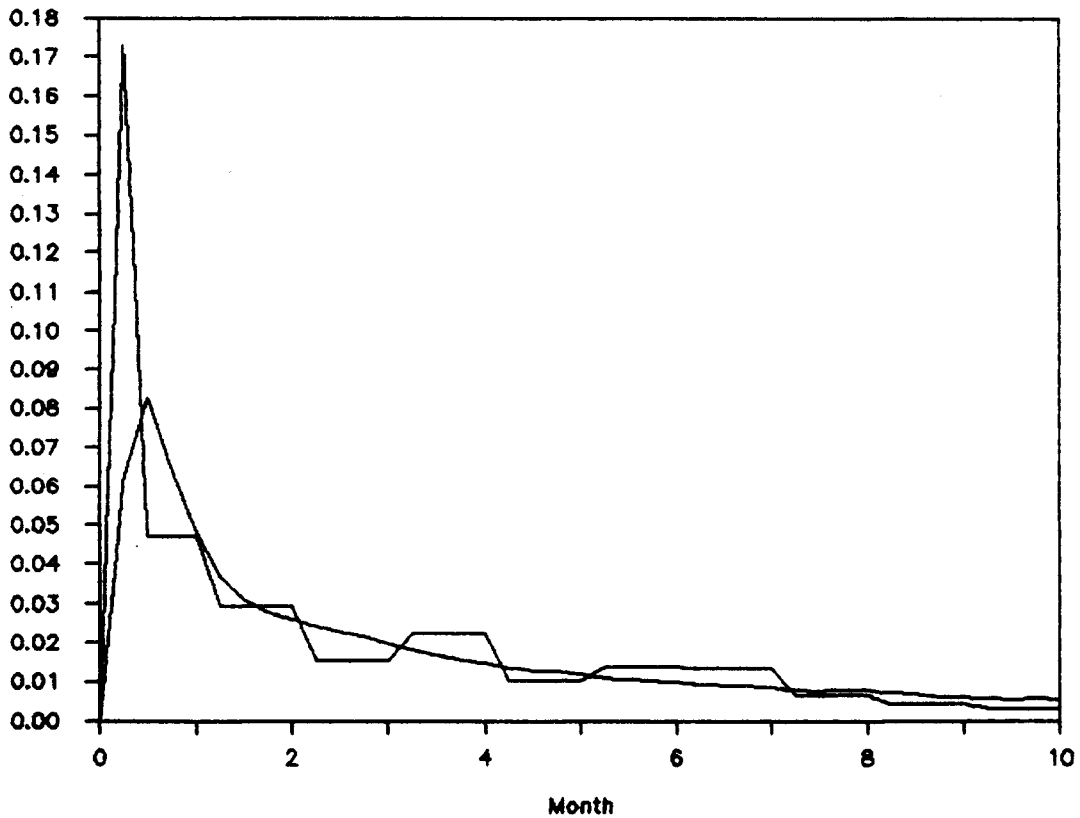
Marginal densities for sample observations

Life table estimate vs. weibull estimate

The plot of the Weibull densities shows a partitioning of
the sample in three groups with rather distinct distribu-
tions of unemployment durations. The largest value for the
heterogeneity parameter implies a strongly skewed density
function with a modal value of less than one month. In con-
trast the mode of durations is about ten months for the
smallest estimate. This results in a unimodal marginal den-
sity function with a mode for a comparatively short duration
of unemployment.

The comparison of the marginal density function from the
model with the corresponding life table estimate reveals a
rather peculiar shape of the empirical distribution. Exits
are frequent for very short durations. After a sharp decline
the frequencies then show a rather flat profile. The esti-
mated marginal distribution function shows a rather good fit
for very short durations and for durations over four months.
In the interval between, the model apparently overestimate
the empirical frequencies.

The shape of the empirical density function for very short
durations reflects to some degree the procedure used for de-
termining unemployment durations. By definition, there are
no spells ending before a duration of a quarter of a month.
After that point, completed durations are measured in multi-
ples of full months while censoring is assumed in the middle
of the month. Thus, the data actually represent a discrete
frequency distribution.

If a mixture model based on the Weibull specification is
fitted to such a distribution, a strong positive duration
dependence is obtained since it allows to represent a
strongly skewed unimodal distribution. If in contrast nega-
tive duration dependence would be assumed, the largest prob-
ability densities would be attributed to the first short du-
ration interval. However, this is not supported by the data

as they are generated since the empirical frequencies are
zero for the first quarter of a month.

This explanation is confirmed by the observation that even
larger values of the Weibull parameter are obtained if more
support points are introduced for the mixing distribution of
the heterogeneity term. The probability density for the
partition with the largest value of the heterogeneity term
becomes even more skewed and concentrated for very short du-
rations. At the same time, in a model with five points of
support, the hump that occurs in the empirical frequency
distribution at about four months of duration is reflected
by an additional partition of the sample.

This implies that the estimate of positive duration depen-
dence may represent an artifact due to inaccuracies of the
data rather than a true causal relationship. If the observed
distribution of durations is truncated from below, it is
difficult to discriminate empirically between a Weibull
model with negative duration dependence and a mixture of
models implying positive duration dependence. The data may
contain not enough information to identify the true form of
duration dependence.

6. Conclusions

Even if unobserved heterogeneity is considered explicitly in
models of unemployment duration it still is rather difficult
to discriminate empirically between true duration dependence
of the hazard rate and the effects of unobserved hetero-
geneity. Economic theory provides only little guidance with
regard to the functional specification of duration depen-
dence as well as to the specification of unobserved hetero-
geneity.

Nevertheless, given the data available, some restrictions
are required for the functional form of duration dependence
and on the distribution of unobserved heterogeneity to
achieve identifiability of the models. If unobserved hetero-
geneity is neglected, a nonparametric specification of dura-
tion dependence becomes feasible. But such models do not
allow causal analyses with regard to duration dependence.
However, they provide an instrument for exploratory data
analyses.

If a functional form is assumed for duration dependence, a
nonparametric approach to unobserved heterogeneity may ap-
pear appropriate since little prior information is available
on the distribution of unobserved heterogeneity and parame-
ter estimates are sensitive with regard to parametric speci-
fications of the mixing distribution. On the other hand,
nonparametric estimates seem to be sensitive with regard to
deficiencies of the data. Since almost no restrictions are
introduced for the mixing distribution, errors in the data
may affect the estimates to a large degree. There is an ob-
vious trade-off between the quality of the data and the
prior information required for estimation.

References

Cox,D. (1972), Regression models and lifetables, in: Journal
    of the Royal Statistical Society, Series B, Vol. 34, pp.
    187-220

Egle,F. (1979), Strukturalisierung der Arbeitslosigkeit und
    Segmentation des Arbeitsmarktes, in: Arbeitsmarktsegmenta-
    tion - Theorie und Therapie im Lichte empirischer Befunde,
    Ch. Brinkmann et. al. (Eds.), Beitrage zur Arbeitsmarkt-
    und Berufsforschung, Vol. 33, Nuremberg

Elbers,C./Ridder,G. (1982), True and spurious duration
    dependence: The identifiability of the proportional hazard
    model, in: Review of Economic Studies, Vol. 49, pp. 403-
    410

Galler,H.P. (1985), Labor supply, labor demand and occupa-
    tional mobility, in: Jahrbuch 1983/84 of the
    Wissenschaftskolleg - Institute for Advanced Study -
    Berlin, , P. Wapnewski (Ed.), Berlin, pp.123-134

Goldfeld,S./Quandt,R. (1972), Nonlinear Methods in Economet-
    rics, Amsterdam

Hanefeld,U. (1984), Das Sozio-Oekonomische Panel - Eine
    Laengsschnittstudie fuer die Bundesrepublik Deutschland,
    in: Vierteljahrshefte zur Wirtschaftsforschung 4/1984,
    pp. 391-406

Hall,J. / Lippman,S. / McCall,J. (1976), Expected utility
    maximizing job search, in: Studies in the economics of
    search, S.Lippman/J.McCall (Eds.), Amsterdam, pp. 133-155

Hausman,J. / Wise,D. (1978), A conditional probit model for
    qualitative choice: Discrete decisions recognizing
    interdependence and heterogeneous preferences, in:
    Econometrica, Vol. 46, pp. 403-427

Heckman,J./ Singer,B. (1984), A method for minimizing the
    impact of distributional assumptions in econometric models
    for duration data, in: Econometrica, Vol. 52, pp. 271-320

- (1986), Longitudinal Analysis of Labour Market Data,
    Cambridge

Hougaard,P. (1984), Life table methods for heterogeneous
    populations: Distributions describing the heterogeneity,
    in: Biometrika, vol. 71, pp. 75-83

Hougaard,P. (1986), Survival models for heterogeneous popu-
    lations derived from stable distributions, in: Biometrika,
    vol. 73, pp. 387-396

Kalbfleisch,J. / Prentice,R. (1980), The Statistical Analy-
    sis of Failure Time Data, New York

Kiefer, N. / Neumann, G. (1979a), Estimation of wage offer
    distributions and reservation wages, in: Studies in the
    Economics of Search, S.Lippman/J.McCall (Eds.), Amsterdam,
    pp. 171-189

- (1979b), An empirical job-search model with a test of the
    constant reservation-wage hypothesis, in: Journal of
    Political Economy, Vol. 87, pp. 89-107

- (1981), Structural and reduced form approaches to analy-
    zing unemployment durations, in: Studies in Labor Markets,
    S.Rosen (Ed.), Chicago/London, pp. 171-185

McCall,J. (1970), Economics of information and job search,
    in: Quarterly Journal of Economics, Vol. 84, pp. 113-126

Mortensen,D. (1970), Job search, the duration of unemploy-
    ment and the Phillips curve, in: American Economic Review,
    Vol. 60, pp. 847-862

Redner,R. / Walker,H. (1984), Mixture Densities, maximum
    likelihood and the EM-algorithm, in: SIAM-Review, Vol. 26,
    pp. 195-239

Solomon,P.J. (1984), Effect of misspecification of regres-
    sion models in the analysis of survival data, in:
    Biometrika, vol. 71, pp. 291-298

Struthers,L.A. / Kalbfleisch,J.D. (1986), Misspecified pro-
    portional hazard models, in: Biometrika, vol. 73, pp. 363-
    369

Trussell,J. / Richards,T. (1985), Correcting for unmeasured
    heterogeneity in hazard models using the Heckman-Singer
    pocedure, in: Sociological Methodology, N.Tuma (Ed.), 1985

What can backward recurrence time data tell us: An application to residential mobility in the U.S.

Nazli Baydar

1. Introduction

Backward recurrence time data contains information on the time of the last ocurrence of an event, or equivalently on the duration spent in the current state. Such data are collected very often, especially by social surveys. Fertility surveys typically provide data on the date of last pregnancy and current nursing status. Some specialized health surveys also collect information on the date of last mensturation and the date of last intercourse that could be used for measuring fecundity. Large scale repetitive cross sectional surveys, like the SIPP or the census in U.S. collect information on duration spent in current residence. These data are the only source of residential mobility data in U.S.

Little is known about the methods of analysis of backward recurrence time data. The main difficulty of such an undertaking lies in the so called sample selection bias. Backward recurrence time data always overrepresents longer durations of stay in the current state. It is impossible to correct for this bias unless one makes explicit assumptions about the nature of the stochastic process that governs the ocurrence of the event of interest. In order to ensure estimability, one has to choose a very simple stochastic process to represent the observed process.

Besides these complications which are due to the sample selection bias, many practical difficulties arise in modelling backward recurrence time data. Most importantly, very little is known about the time path of the covariates of the process of interest. In most cases, only the current values of the covariates are known. If these covariates are not time independent, their utility is very limited.

In spite of its shortcomings, the backward recurrence time data allows one to infer some characteristics of the underlying process. I investigate the possibility of using the theory of renewal processes for the univariate and multivariate analysis of backward recurrence time data on residential mobility. Derivation of the essentials of the method is general. However application of the method to a particular process is not always straightforward. Each application requires the modification of the basic method to match the specific features of a process. For this reason, I will restrict my attention to residential mobility.

Section 2 contains a literature review and an introduction to the elementary methods of analyzing backward recurrence times. Estimation of the duration dependence of the hazard rate from backward recurrence time data draws heavily on the theory of renewal processes. This method and its extention to the estimation of multivariate models of the hazard rate are presented in sections 3 and 4. Multivariate hazard models for backward recurrence time data can only be estimated by using some numerical techniques of analysis. These techniques are explained in section 5. Section 6 presents an application to U.S. census data on residential mobility.

2. Elementary methods of analyzing backward recurrence time data

The properties of backward recurrence times have been known among mathematical statisticians since the 1960's. Most of the early work on the theory of backward recurrence times has been limited to the simple case of a Poisson process. A Poisson process is characterized by a constant hazard and an exponential density. Since this stochastic process has no memory, the derivation of the properties of the backward recurrence times is particularly simple.

Drawing upon Feller's (1966) work, social scientists in the 1970's began to make use of the theory of backward recurrence times to analyze data available from many surveys. Most applications were on open ended birth interval data or on current nursing status data (Sheps et al., 1970; Menken and Sheps, 1970). More recently, Ginsberg (1983) and Allison (1985) applied the Poisson model to migration. Sorensen (1977) generalized Feller's work and showed that the magnitude of the bias in observed mean duration of stay that is calculated from backward recurrence time data is related to the mean and the variance of the density of the true process.

Cox (1962) proved that for any renewal process, the distribution of the backward recurrence times can be written in terms of the renewal function and the survival function of the true process. The main problem of applying such statistical theory to the analysis of social processes is that individuals experiencing events of a given kind behave differently from 'light bulbs' experiencing 'renewals'. Most importantly, individuals experience a variety of interrelated events in different domains. The renewal model gives an oversimplified representation of such interactions.

Even with such oversimplified models, analysis of backward recurrence times may be quite difficult. Backward recurrence time data cannot be a substitute for complete event histories. However, backward recurrence time data may still be of value when it is impossible to collect complete histories either due to recollection problems or due to financial constraints.

Figure 1 presents a schematic representation of backward recurrence time data. I denote the time of the survey by t_s, and the reported backward recurrence time by u. We know the current age (i.e. u) of the individuals who never experienced the event of interest. In some sense, for these individuals, we know the complete event history. In order to estimate hazard models from such data, two steps must be taken: (1) some simplifying assumptions must be made about the nature of the true process; (2) the density of the backward recurrence time, $f^*(u)$, must be written in terms of the known characteristics of the underlying stochastic process. The relevant characteristics of the underlying stochastic process are the hazard, density and the distribution functions. Let h(t) denote the hazard rate for time t. From the knowledge of the hazard rate one can derive the density, f(t), distribution function, F(t), and the survival function S(t):

$$h(t) = \frac{f(t)}{1 - F(t)} \qquad (1)$$

$$dF(t) = f(t)\ dt \qquad (2)$$

$$S(t) = 1 - F(t) = exp\ \{\ -\int_0^t h(z)\ dz\ \} \qquad (3)$$

We wish to write the density of the backward recurrence time in terms of these functions. The simplest renewal process is the Poisson process. The density of the backward recurrence time of this process is well known:

$$f^*(u) = a \exp(-au) \qquad\qquad (4)$$

where a is the hazard rate of the process. It can be easily shown that the mean duration of the spell that surrounds the survey is twice the mean length of the spells dictated by the process. This phenomenon is known as the length bias in sampling or the waiting time paradox.

Expression (4) is intuitively appealing and it provides a simple basis to understand the derivation of the backward recurrence time density for more complex renewal processes. For the individuals who experienced an event u time units ago, we know that they survived experiencing another event between $t_s - u$ and t_s. The exponential term represents this survival probability [cf. equation (3)]. For these individuals we also know that they experienced an event at time $t_s - u$. For a Poisson process this rate is independent of past history or time. The hazard rate (a) contributes this information to the density given in (4).

The expression for the density of u is necessary but not sufficient to estimate models of backward recurrence times. In order to estimate the hazard rate (a) under the Poisson process, one needs to derive an expression that represents each individuals contribution to the likelihood. The individuals who never experienced an event before contribute to the likelihood the probability of surviving for u time units without an event.

In order to complete the derivation of the likelihood one also needs to include a term that represents the probability of censoring (probability of being removed from observation) before t_s. Hoem (1969) derived expressions for transition probabilities of a Markov chain when some individuals are censored before they could report their event histories. If censoring is not selective by the characteristics of the individuals that

are also relevant to the process of interest, it is relatively simple to correct the likelihood. Let the hazard functions pertaining to the events that lead to censoring be $c_i(t)$. Let there be I such events (for example emigration, death, institutionalization etc.). If these rates operate independently from each other (which is often an implausible assumption, especially for residential mobility), the rate of censoring is the sum :

$$c_+(t) = \sum_I c_i(t).$$ (5)

The probability of surviving censoring until time t_e is:

$$S_e(t_e) = exp \ \{-\int_0^{t_e} c_+(t) \ dt\}$$ (6)

The contribution of each individual to the likelihood may now be written as:

$$L = a^d \ exp(-au) \ S_e(t_e)$$ (7)

where d is an indicator variable that is 1 if the individual has ever experienced an event.

Note that this expression is true only under the following assumptions: (1) the hazard rate is independent of time; (2) the hazard rate is independent of individual characteristics; (3) different censoring mechanisms operate independenty from each other; (4) the censoring mechanisms operate independently from the process of interest. Although the last two assumptions are routinely made in the analysis of event histories, they are by no means plausible. There is some recent work to incorporate the dependencies between risks, however review of these developments is beyond the scope of this paper. The first two assumptions and the ways to relax them for modelling backward recurrence time data is the subject of the next two section.

3. Can we model the time dependency of the hazard rate if we have backward recurrence time data only?

The reason why most of the applications of the theory of backward recurrence times adopt the Poisson model is because even the simplest forms of time dependence of the hazard results in substantial complexity in the method. A time dependency implies that the process has some memory: the probability of ocurrence of the next event depends on when the previous event has ocurred. If the relevant history is not known, the probability of experiencing an event u time units before the survey [f*(u)] depends on all possible time paths of the process before time $t_s - u$. When the process has no memory the time path of the process becomes an irrelevant concern.

Models of time dependent hazard functions are estimable from backward recurrence time data if one assumes that the underlying process is a renewal process. This assumption implies that subsequent events of interest are indistinguishable. Undoubtedly, this assumption leads to an oversimplified view of an individual's event history.

In the analysis of fertility histories with complete data, the difficulty of assuming independence between spells is overcome by analyzing parity specific birth intervals. In the analysis of geographic mobility histories or employment histories, the dependencies between subsequent spells are often ignored even if complete event histories are available. The problem of modelling multiple spells is difficult not only methodologically but also conceptually, since there is little theoretical guidance about the nature of the dependencies between spells. When the data is limited to backward recurrence times, the assumption of independent

identically distributed spells is indispensable for the development of a tractable methodology.

The density of backward recurrence times under the renewal model consists of two components, quite similar to the expression given in (4). The first component is the probability of experiencing an event u time units before the survey, and the second component is the probability of surviving without another event until the survey. The derivation of the second term presents no difficulties. The first term is the renewal density. The renewal density at any given time depends on the density of the spells in a complex way. The derivation and the properties of the renewal density is discussed in detail in standard text books of stochastic processes (e.g. Cox, 1962; Karlin and Taylor, 1975). Below, I summarize the derivation of an analytical expression for the renewal density (for a more detailed discussion see, Baydar and White, 1986).

Let $M(t)$ denote the renewal function. This function is related to the distribution function $F(t)$ by the following convolution equation:

$$M(t) = F(t) + \int_{0}^{t} M(t-z) \, dF(z) \, dz \qquad (8)$$

The only way to evaluate the renewal function is by taking the Laplace transform of the distribution function:

$$M_1 = \frac{F_1}{1 - F_1}, \qquad (9)$$

where M_1 and F_1 represent the Laplace transforms of the renewal function and the waiting time distribution function. The Laplace transform of the

renewal function can then be inverted to obtain the renewal density. The Laplace transform is defined as:

$$F_1 = \int_0^\infty exp(-1t) \; dF(t) \qquad (10)$$

where 1 is the Laplace transform parameter. Following sections will refer to the function under the integration as L(t).

The process of deriving an expression for the renewal density is very tedious. That is why most textbook problems deal with the limiting case. As time is further removed from the origin, the renewal function approaches a constant that is equal to the inverse of the mean length of the spells. For the analysis of backward recurrence time data for individuals, the assumption that the limiting solution applies to all individuals is not plausible. This assumption implies that after the start of the process of interest, each individual experienced sufficiently long durations of exposure to the risk of experiencing an event. Here, 'sufficiently long' should be interpreted relative to the mean length of the spells. The duration required for the convergence to the limiting value is proportional to the mean waiting time and inversely proportional to its variance. I found that in most practical applications an exposure of twice the length of the mean waiting time can be considered long enough.

Once the renewal function is known, one can use the formula for the backward recurrence times given first by Cox (1962):

$$Pr\{U \; > \; u\} = \int_0^{t_a -u} [1 - F(t_a -z)] \; dM(z) + [1 - F(t_a)] \qquad (11)$$

The backward recurrence time density that we need to insert in the likelihood function is:

$$p(u) = m(t_s - u) [1 - F(u)]$$ (12)

The likelihood function comparable to the one given in (7) is:

$$L = m(t_s - u)^d S(u) S_s(t_s)$$ (13)

where d is defined similarly.

The main difficulty in deriving analytical expressions for the renewal density lies in the Laplace transformation and the inversion of this transform. Some functions that are commonly used to represent the time dependency of the hazard do not have tractable Laplace transforms (F_1) and many such functions result in expressions for M_1 that cannot be inverted analytically. I discuss some numerical procedures for the evaluation and inversion of the Laplace transforms in section 5.

The density of the backward recurrence times under a renewal model other than the Poisson model is a function of the time of the survey, t_s [see equation (12)]. t_s is measured relative to the time origin (t_o) of the process. In the case of residential mobility, time origin is the point in time after which an individual with a given set of characteristics, which are relevant to the process, is at risk of giving residential mobility decisions. Consider, for example, an analysis of the residential mobility of the retired individuals classified by marital status. The time-origin for a retired widower is equal to the time of retirement or the time of becomong a widower, whichever event ocurred later. If the retirement ocurred later, t_s is equal to the duration spent in retired state.

4. Can we model the effects of covariates on the hazard rate using backward recurrence time data only?

Estimation of the multivariate models of the hazard rate from information on the backward recurrence times depends, first of all, on whether we can derive an expression for the density, $f^*(u)$. Once this expression is found, the question arises whether that expression can be numerically evaluated. In the previous section I showed that $f^*(u)$ is a product of two terms, the renewal density and the survival function. It is the derivation of the renewal density that is considerably more difficult under the multivariate model.

The renewal function at any given time t depends on the entire time path of the process since the time origin. If the covariates are time dependent, the time path of the process becomes dependent upon the time path of the covariates. The time path of the covariates is almost never known from the surveys that provide backward recurrence time data. Therefore we need to treat the covariates as fixed after a certain time point. I shall assume that for every individual we know a point in time (t_0) after which all relevant covariates have remained fixed. Figure 2 shows how t_0 can be identified. Consider a familiar parametric multivariate model for the hazard rate where the covariates are fixed after t_0:

$$h(t) = h_p(t) \, exp(X\beta) \tag{14}$$

where $h_p(t)$ represents the parametric model of time dependence of the hazard rate. In what follows a p subscript will indicate that the function concerned is determined by the $h_p(t)$ only, and not by the covariate vector.

$h_p(t)$ can also be regarded as a baseline hazard. The survival function can be written down immediately, following the definition in (3):

$$S(t) = exp \{ -exp(X\beta) \int_{t_o}^{t} h_p(z) \, dz \}$$

$$= S_p(t)^{exp(X\beta)} \tag{15}$$

In order to derive the renewal function, we need to repeat the steps described in section 3. That is, we need to take the Laplace transform of the distribution function. Let 1 be the Laplace transform parameter. Then F_1 is:

$$F_1 = exp(X\beta) \int_{0}^{\infty} exp(-1t) \, S_p(t)^{exp(X\beta)-1} \, f_p(t) \, dt. \tag{16}$$

It is aparrent from equation (16) that there is no analytical solution to the Laplace transform of the distribution function. This is true for all but one possible function for $h_p(t)$. If the baseline hazard is constant the expression (16) can be integrated analytically. If F_1 can be analytically evaluated, M_1 and $m(t)$ can be evaluated as shown in (9). One can easily show that the complete likelihood function for each individual under the multivariate constant rate model is:

$$L = [a \ exp(X\beta)]^d exp(-au) \, S_a(t_a) \tag{17}$$

If one wants to estimate a multivariate model of a time dependent hazard function using backward recurrence time data, the evaluation of the $f^*(u)$ has to be done numerically.

5. Is it possible to evaluate the renewal function numerically?

The difficulty of evaluating the renewal function numerically is twofold: both the evaluation of the Laplace transform (F_1) and the inversion of the Laplace transform to evaluate m(t) are numerically sensitive operations.

The numerical evaluation of the Laplace transform given in (15) is very difficult. Since the Laplace transform parameter (l) is a complex number, its sine and cosine components result in very large fluctuations when t is close to zero. Figure 3 presents the L(t) function whose integral is the Laplace transform of the Erlangian density with shape parameter 2. It can be seen that both the real and the imaginary parts of L(t) fluctuate when t is small. However, nearly all numerical integration algorithms assume reasonable smoothness of the function to be integrated. When this assumption does not hold, the F_1 values become very sensitive to the choice of the parameters of the numerical algorithm.

The second source of difficulty lies in the numerical inversion of the Laplace transform of the renewal function. The numerical inversion of the Laplace transform is a difficult problem known to many branches of the engineering sciences (for an overview of such problems see Bellman et al., 1966). This is mainly because of the instability of the Laplace transformation: a small perturbation in the original function may have a large impact on its Laplace transform. For a reasonably accurate numerical inversion, the Laplace transform itself has to be evaluated to a considerable precision. This is not possible unless one uses extra-ordinarily accurate numerical algorithms of integration.

In order to improve the efficiency of the numerical techniques, a few strategies can be followed:

1) The Laplace transform integral can be integrated by parts before the use of the numerical algorithm. If the complex component is integrated and the real component is differentiated, one can obtain a smoother function to be input into the numerical integration. The following equation can be used for integration by parts:

$$F_1 = (1/l) \ exp(X\beta) \ [\lim_{t \to 0} f_\bullet (t)]$$

$$+ \ (1/l) \ exp(X\beta) \ \int_0^\infty exp(-lt) \ [\{exp(X\beta)-1\} \ S_\bullet (t)^{\{exp(X\beta)-2\}} \{f_\bullet (t)\}^2$$

$$+ \ S_\bullet (t)^{\{exp(X\beta)-1\}} \{df_\bullet (t)/dt\}] \ dt \qquad (17)$$

2) The renewal function for the individuals who have experienced a long enough exposure is very close to its limiting value, the inverse of the mean length of the spells. The contribution of those individuals to the likelihood can be approximated without the numerical evaluation of the Laplace transform. The mean length of the spells is:

$$\mu = \int_0^\infty S_\bullet (t)^{exp(X\beta)} \ dt. \qquad (18)$$

6. An application of the univariate and multivariate hazard models to the backward recurrence time data on residential mobility

I applied the methods described in the previous sections to U.S. residential mobility data from the 1980 census. U.S. census provides backward recurrence time data on residential mobility of the heads of the households (householders). For each householder we know the year of arrival to the current residence unit. The public-use microdata (PUMS) file provides a random sample of all householders. This data set also provides some socio-demographic covariates. Unfortunately, the dates at which an individual attained the current covariate values are not known. Therefore, I had to make certain simplifying assumptions to determine the time origin of the process for each individual. Table 1 summarizes where the time origin is set for each covariate. Age 18 is assumed to be the earliest age at which individuals can make their own mobility decisions. This assumption is rather arbitrary, however the results of the analysis are not very sensitive to small shifts in the time-origin (Baydar and White, 1986). I restrict the analysis to a 5 percent sample of the PUMS file and to those householders who are born after 1944 (aged under 35 at the time of the census). This gives me a sample of 1296 householders.

Application of the univariate backward recurrence time models is fairly straightforward once an expression for $f^*(u)$ is found. An exploratory analysis revealed that a non-monotonic hazard function is appropriate to represent the duration dependence of residential mobility for this cohort (Baydar and White, 1986). I adopted a mixture model for representing the time dependence of the hazard. The underlying density of

this model is a mixture of an exponential density and an Erlangian density with shape parameter 2. Let a and b be the parameters of the exponential and the Erlangian density respectively; and let δ_1 and δ_2 be the mixing coefficients. The density of the failure times can be written as:

$$f(t) = \delta_1 a^2 t \exp(-at) + \delta_2 b \exp(-bt). \qquad (18)$$

In order to construct the full likelihood as described in (13), one also needs to specify the censoring mechanisms. In this case censoring may occur because of either death or emigration. In the absence of any reliable emigration statistics in U.S., it is not possible to estimate probabilities of surviving emigration. However, ignoring emigration is not expected to have a large impact on the estimates since its magnitude is very small. We use age, sex and race specific life tables of 1980 (Bureau of the Census, 1981) to correct the likelihood function for censoring due to mortality.

The mixture model given in (18) can be interpreted as a simple model of unobserved heterogeneity. Such mobility models with only two latent classes are also known as mover-stayer models (Blumen et al., 1955; Morrison, 1971). The formulation in (18) is consistent with the hypothesis that the movers experience increasing risks of residential mobility over duration, and the stayers experience a low and constant risk. Figure 4 shows these components of the residential mobility hazard for a subsample of the whole cohort that consists of homeowners.

In order to do a univariate analysis of these residential mobility data I subsampled the observations according to a selected set of dichotomised covariates. The detailed results of this analysis are reported in a previous paper (Baydar and White, 1986). Table 2 summarizes the findings. Subsampling by homeownership and the presence of school-aged

children result in largest residential mobility differentials. Since the univariate analysis is done by estimating models independently for each subsample it also provides some clues to the non-proportionality of the hazards over the categories of some covariates. The two occupation covariates result in non-proportional hazards for professionals vs. others and for production occupations vs. others. This feature is difficult to incorporate in a multivariate analysis.

The two-sample t-statistics show that the differences in the hazard rates of the movers between the two categories of a variable are often statistically insignificant, the only exception being the homeownership variable. This finding implies that the individuals who are dissatisfied with their current residence (the 'movers') tend to move irrespective of their household or individual characteristics. Such characteristics result in differentials of residential mobility only among the group of 'stayers', i.e. those who are currently satisfied with their residence.

The multivariate analysis of the residential mobility of this sample is not yet completed. The results that I can present here are those of a multivariate constant rate model and a multivariate renewal model where the limiting solution for the renewal density is adopted. The numerical procedure to estimate the finite time renewal density is still being developed.

Table 3 presents the results of the two multivariate models. The estimated duration dependence of the multivariate renewal model indicates that the hazard is almost constant over the duration of residence (the proportion of movers is very small). Allison (1985), after applying the limiting renewal density for a Gompertz model, also found that the

estimated duration dependence was not significantly different from the constant hazard model. This finding is probably an artifact of the assumption that the limiting renewal density applies to all individuals irrespective of their actual exposure time. The example presented in Appendix A shows that for the univariate case, if the limiting solution to the renewal density is adopted, the non-monotonic pattern of the hazard rate disappears.

According to the multivariate analysis most of the covariates do not have significant effects on the residential mobility rate. There may be two reasons for this: (1) the small size of the sample or (2) the misspecification of the duration dependence of residential mobility. The sample size problem can easily be overcome by increasing the size of the sample that is drawn from PUMS file. It is difficult to see how a constant hazard specification will affect the inferences about the covariate effects when there is strong evidence from the univariate analysis that the residential mobility hazard is non-monotonic.

The two models presented in Table 3 result in consistent inferences about the magnitude, direction and the significance of the covariate effects. Homeownership and the presence of school aged children have large negative effects on the residential mobility rate. When controlled for the presence of school-aged children the effect of being married and the effect of living in a family household disappear. The effect of being black is significant although the univariate analysis does not reveal strong racial differentials. The occupational differentials are not significant under the multivariate model, probably because the hazards for different occupational categories are not proportional, as remarked earlier.

7. Conclusions

Backward recurrence time data is a very poor source of information about individuals' event histories. However, when complete event histories are not available for analysis, one can make use of backward recurrence time data. Renewal theory proves to be particularly useful for the derivation of the density of the backward recurrence times. Once this density is obtained, parameters of a model can be estimated by maximum likelihood techniques.

The use of the renewal theory for estimation requires the following assumptions to be met: (1) the event that is modelled is a repetitive event and subsequent events of the same kind are indistinguishable; (2) for each individual there exists a time point after which the covariates of the stochastic process have remained fixed until the time of the survey; (3) the censoring mechanisms and the corresponding probabilities of surviving censoring are known; (4) the censoring probabilities and the probabilities of experiencing the event of interest are independent.

A very limited set of hypotheses can be tested under these assumptions. The interdependencies among various kinds of events and the effects of past history on the future behavior cannot be investigated using backward recurrence time data. However this method is superior to the regression or logit analysis of backward recurrence times. Besides the questions concerning their technical validity, these more conventional methods do not allow one to infer the duration dependency of the hazard rate or to correct the estimates for length biased sampling.

Table 1. Time-origins for the covariates

| Covariate	Category	Time-origin
Marital status	Never-married	18
	Ever-married	Age at first marriage
Homeownership	Homeowners	25
	Tenants	18
Household type	Family households	Age at first marriage
Non-family hh.	18	
School-aged children	Children present	Age when the oldest child reaches age 6
	Children absent	18
Occupation	Professionals	22
Others	18	
Other covariates		18

Table 2. Results of the univariate analysis of backward recurrence time data on residential mobility (* = significant at .01)

Variable Parameters	Category 1	Category 2	Two sample t-statistic
Gender	Female	Male	
a	2.2023	2.3561	1.36
b	.2307	.2706	2.82*
δ_1	.6127	.4984	3.81*
Race	Black	Non-black	
a	2.1643	2.3484	1.06
b	.2129	.2715	3.18*
δ_1	.4468	.5299	2.30
Marital status	Ever-married	Never-married	
a	2.3585	2.4550	.91
b	.2135	.2656	4.40*
δ_1	.4553	.7170	9.89*
Homeownership	Homeowners	Tenants	
a	1.6737	2.3914	7.28*
b	.1468	.2930	10.34*
δ_1	.3545	.6798	13.28*
Household type	Family hh.	Non-family hh.	
a	2.4780	2.3256	1.46
b	.2062	.2360	3.14*
δ_1	.4634	.6869	9.12*
School children	Present	Absent	
a	2.1795	2.3556	1.00
b	.1989	.2960	5.61*
δ_1	.0289	.4852	38.39*
Occupation	Production	Other	
a	2.3244	2.3226	.02
b	.2179	.2958	5.51*
δ_1	.6152	.4509	6.71*
Occupation	Professional	Other	
a	2.0115	2.3248	2.25
b	.2873	.2392	3.10*
δ_1	.3363	.5932	11.49*
Education	High-school +	High-school)	
a	2.2926	2.3777	.43
b	.2721	.2353	1.17
δ_1	.4655	.6483	1.97

Table 3. Results of two multivariate models of backward recurrence time data on residential mobility; 1980 U.S. census, householders under 35.

	Constant hazard model		Erlangian and exponential mixture model	
	Parameter	t-statistic	Parameter	t-statistic
Hazard parameters				
a	.6219	8.91	2.4463	24.10
b			.6080	9.34
δ_1			.0177	7.36
Covariates				
Homeownership	-.7092	8.29	-.6641	7.85
Family household	-.0093	.08	.0456	.40
Professional occup.	-.1157	1.22	-.0977	1.05
Production occup.	-.1546	1.70	-.1453	1.61
Female	-.0141	.14	-.0238	.24
Black	-.4620	3.84	-.4557	3.79
Hispanic	.1114	.56	.1638	.85
School-aged child.	-.3791	3.87	-.3885	3.99
Ever-married	-.1036	.94	-.1508	1.36
High school educ.	-.0101	.12	.0029	.04

Figure 1. Backward recurrence time data
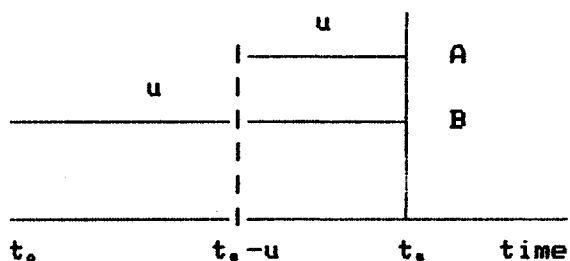
Note: Individual a experienced an event at time $t_* - u$; individual B did not experience any events. For individual B, u is equal to the duration since the start of the process (age).

Figure 2. Identification of the time origin of the process for a

hypothetical individual
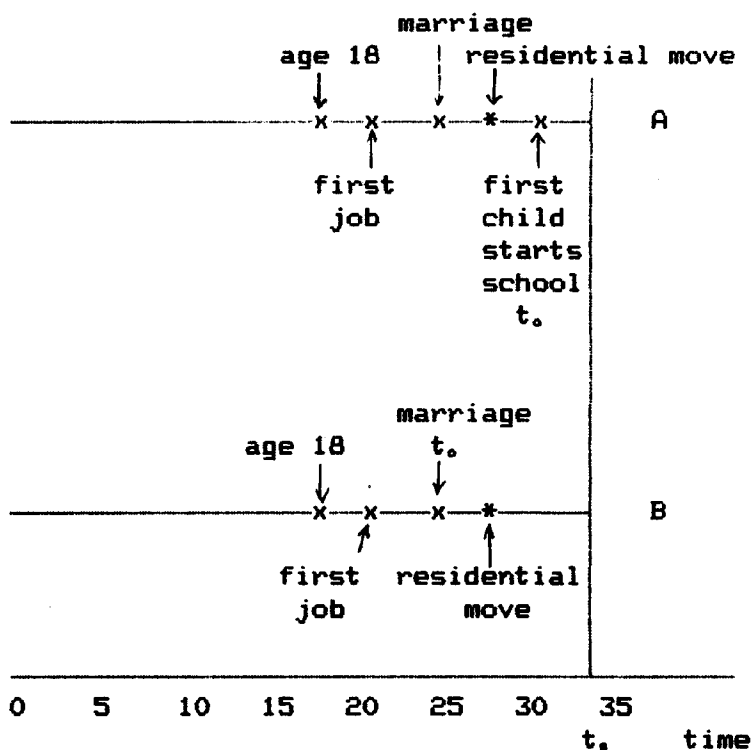
Note: Time origin (t_*) is the time of the last change in the relevant covariate set. Covariates of the individual A change after a residential move. Therefore this individual contributes a survival term to the likelihood. Individual B contributes a renewal term in addition to the survival term since the last change in the covariate set occurred before the residential move.

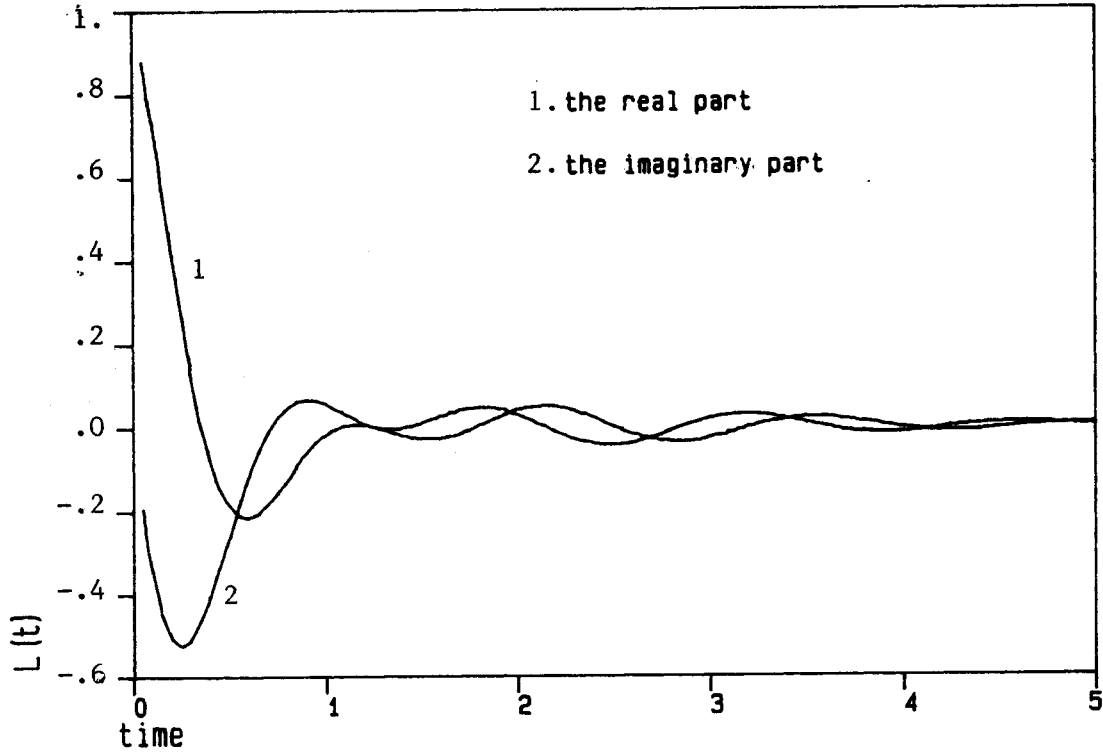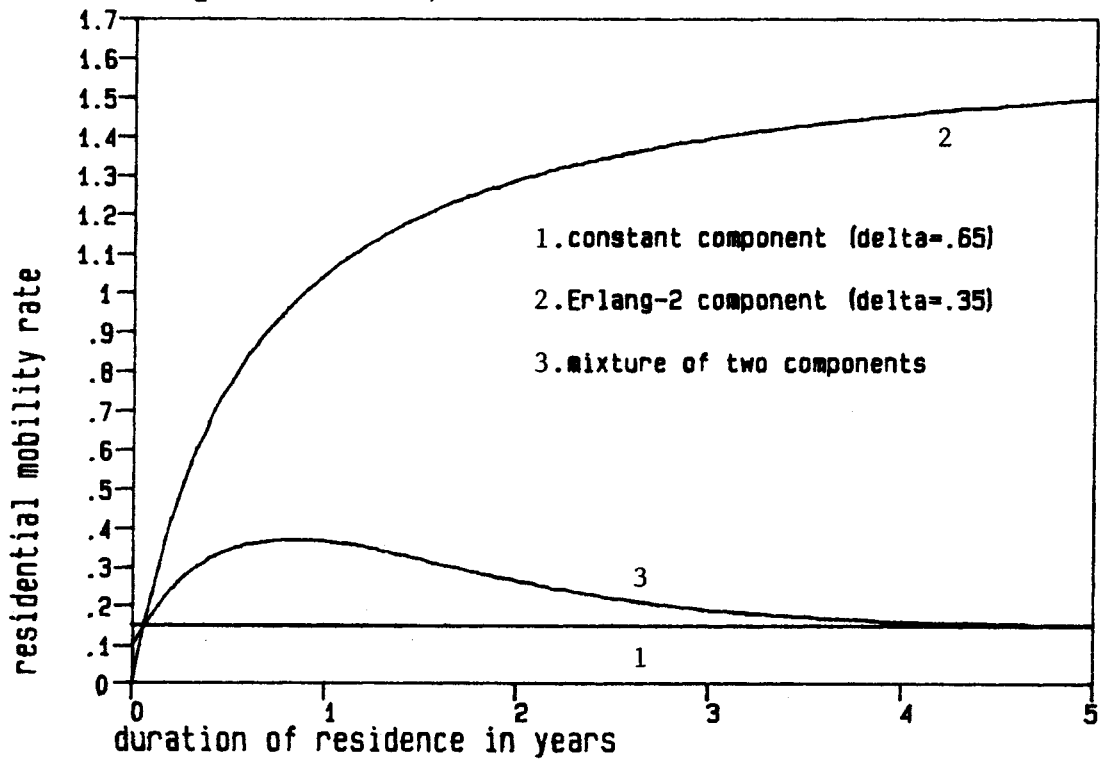Figure 3. L(t) function for an Erlang-2 density (l=.44 + 4.4i)

1.the real part

2.the imaginary part

Figure 4. Components of the mixture model

1.constant component (delta=.65)

2.Erlang-2 component (delta=.35)

3.mixture of two components

References

Allison, P. D. (1985) Survival analysis of backward recurrence times, JASA, 80, 315-322.

Baydar, N. and White, M. (1986) Residential mobility as a renewal process: mover-stayer models · for backward recurrence time data, to be published.

Bellman, R. E., Kalaba, R. E. and Lockett, J. A. (1966) Numerical inversion of the Laplace transform: applications to biology, economics engineering and Physics, New York: American Elsevier Publishing Co.

Blumen, I., Kogan, M. and McCarthy, P. J. (1955) The industrial mobility of labor as a probability process, Cornell Studies in Industrial and Labor Relations, no. 6, Ithaca, Cornell University.

Bureau of the Census (1981) Statistical Abstract of the United States, 1981, Washington, D.C.: U.S. Department of Commerce.

Cox, D. R. (1962) Renewal Theory, London: Methuen and Co., Ltd.

Feller, W. (1966) An introduction to the probability theory and its applications, Vol. 2, New York: Wiley.

Ginsberg, R. B. (1983) Moving in a given year: a study in research design and data analysis, Tijdschrift voor Economische en Sociale Geografie, 74, 253-266.

Hoem, J. (1969) Purged and partial Markov chains, Scandinavisk Aktuarietidskrift, 152, 147-155.

Karlin, S. and Taylor, H. M. (1975) A first course in stochastic processes, New York: Academic Press.

Menken, J. and Sheps, M. C. (1970) On relationship between longitudinal characteristics and cross-sectional data, American Journal of Public Health, 60, 1506-1511.

Morrison, P. (1971) Chronic movers and the future redistribution of population: a longitudinal analysis, Demography, 8, 171-184.

Sheps, M. C., Menken, J. A., Ridley, J. C. and Lingner, J. W. (1970) Truncation effect in closed and open birth interval data, JASA, 65, 678-693.

Sorensen, A. B. (1977) Estimating rates from retrospective questions, in: Heise, D. R. (ed.), Sociological Methodology 1977, San Francisco: Jossey-Bass, 209-223.

Appendix A. A comparison of the finite time and the limiting estimates of duration dependence

The density of the backward recurrence time [f*(u)] is a function of the time of the survey (t_s) measured relative to the time-origin (t_o) of the process. This dependence disappears as t_s is further removed from t_o. When t_s is sufficiently large, the renewal density converges to a constant that is equal to the inverse of the mean waiting time of the process. In this case, the backward recurrence time density can be evaluated without evaluating the renewal density.

In many sociological and demographic applications of the backward recurrence time modelling, the assumption that each individual had a long enough exposure to the risk of experiencing a particular event is not appropriate. It is therefore incorrect to adopt the limiting solution to the renewal density. The following example shows that the impact of this error may be substantial.

I estimated the same model of the duration dependence of residential mobility, once adopting the finite time solution and next adopting the limiting solution to the renewal density. The duration dependence model is a mixture model: the mixture of an Erlang-2 and exponential densities. The sample consists of ever-married householders at the 1980 U.S. census, who were under age 35 at the time of the census. Figure A.1 presents the hazard rates of residential mobility estimated by the finite time and the limiting solutions. The finite time solution results in a clearly non-monotonic pattern of the hazard rate. The limiting solution to the renewal density leads to erroneous conclusion that the hazard rate is constant over

duration of residence. The impact of the misspecification of the renewal density in this example is very large even though the mean waiting time of this particular parocess is fairly small (approximately 3 years), and a large proportion of the sample has experienced longer exposures after the date of first marriage (t_o).

Figure A.1. Comparison of finite time and limiting estimates
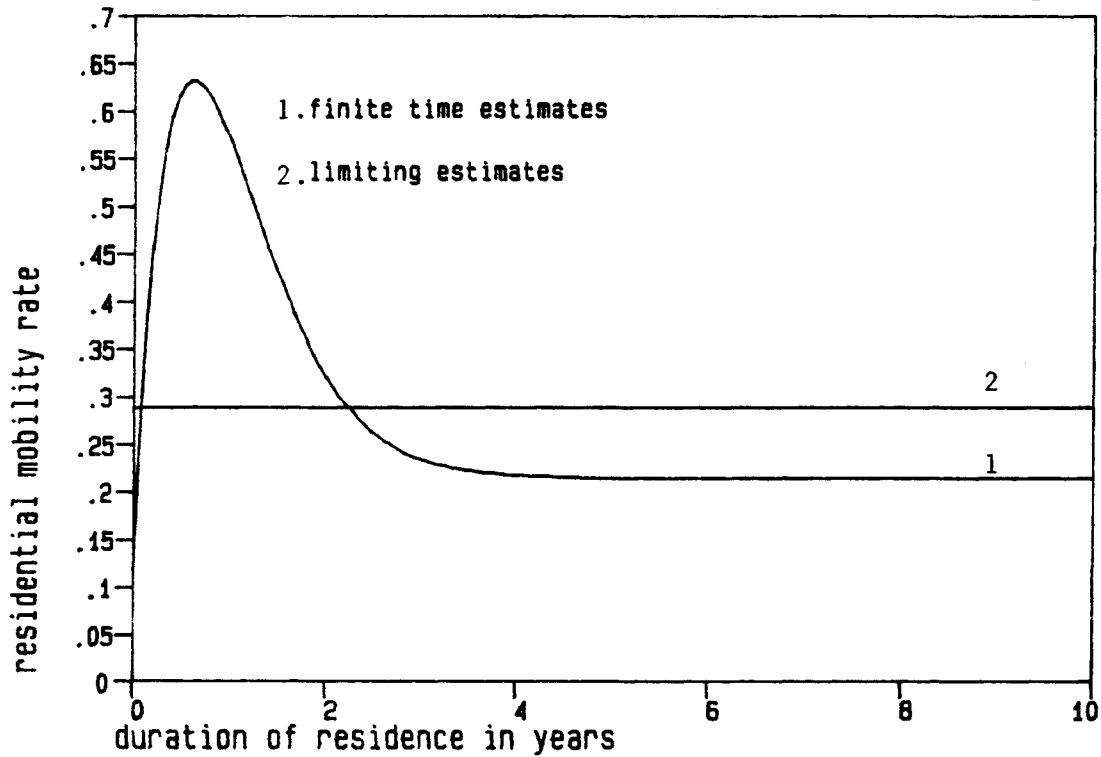
Testing against misspecification in parametric rate models

Gerhard Arminger

1. Misspecification of models

A crucial issue in the interpretation of results from fitting
statistical models to empirical data is the underlying but rarely
outspoken assumption that the class of models fitted to the data
contains the true model of the random mechanism generating the
data. If this is not the case, it may easily happen that the
parameters estimated from the wrong model are heavily biased and
lead to wrong conclusions about the nature of relationships and
causal mechanism. Main problems in the context of rate models
- like in any other regression type model such as generalized
linear models or covariance structure models - are the correct
functional form of the rate, especially the dependence on the
duration time and on relevant covariates including functions of
explanatory variables and interactions. Since the interpretation
of results from model fitting depends completely on the
assumption that a correct model has been estimated, general
tests against misspecification without knowing the form of
alternative hypotheses are necessary. Although such tests have
been developed in econometrics (Cf. the seminal work of Haus-
man, 1978; White, 1980a, 1980b, 1981, 1982; White and Domowitz
1984) in the context of linear and nonlinear regression models
as well as for general maximum likelihood estimation procedures,
these tests have hardly been applied even in the econometric
literature. One of the main obstacles to their more frequent
use may be the reliance on asymptotic theory requesting large
sample sizes for the tests to be useful which may be hard to obtain
in econometric data sets. However, large samples are quite common
in fields such as sociology, psychology, demography and
epidemiology, and here the tests should certainly be applied
to prevent researchers from drawing wrong conclusions from
their estimates. Since the focus in this paper is on parametric
rate models which are usually estimated by maximum likelihood

methods the tests of misspecification proposed by White (1982) which have been derived from ML estimation under misspecification will be applied. Weights for tests of the Hausman type (Hausman 1978) against inconsistency of parameters will be chosen by considering deviance increments in rate models developed in a companion paper (Arminger, 1986).

2. ML estimation of misspecified models

Before applying the matrix information test of White (1982) and Hausman type tests to rate models, the key results of White (1982) are briefly reviewed. For the results to hold true, the regularity assumptions of White (1982) have to be fulfilled and differentiation under the integral sign is allowed. Let $U_i \sim 1 \times M$, $i=1,\ldots,n$ be independently identically distributed random vectors with true density $g(u)$. The model for the random vector U supposed by the researcher is a parameterized family $f(u,\theta)$ of densities dominated by the same measure ν as $g(u)$ which we assume to be the Lebesgue measure $\lambda(u)$. θ is a $p \times 1$ vector. If the model is correctly specified, the family $f(u,\theta)$ contains as an element $g(u) = f(u,\theta_o)$.

A quasi maximum likelihood estimator $\bar{\theta}_n$ (QMLE) maximizes for given n the quasi loglikelihood

$$l_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ln f(u_i,\theta) \qquad (2.1)$$

If $g(u) = f(u,\theta_o)$, $l_n(\theta)$ is the log likelihood of the sample. Under suitable regularity assumptions $l_n(\theta)$ converges almost surely against the expected value of $\ln f(u,\theta)$ that is $\int \ln f(u,\theta)g(u)d\lambda(u)$. If $E_g[\ln f(u,\theta)]$ has a unique maximum

at θ_*, the QMLE $\bar{\theta}_n$ converges almost surely against θ_*.
The last condition is identical to the condition that the
Kullback Leibler information criterion (KLIC) which measures
the discrepancy between two densities

$$I(g;f,\theta) = \int \ln \frac{g(u)}{f(u,\theta)} g(u)d\lambda(u) \qquad (2.2)$$

has a unique minimum at θ_*. θ_* is the value of θ which
minimizes the discrepancy between $g(u)$ and $f(u,\theta)$. Hence, $\bar{\theta}_n$
may also be called the minimum ignorance estimator. The KLIC
$I(g;f,\theta) \geq 0$ and is equal to 0 only for $g(u) = f(u,\theta)$ almost
everywhere. Hence, if $g(u) = f(u,\theta_o)$, the vector θ_* that minimizes
$I(g;f,\theta)$ must be equal to θ_o and θ_n converges to θ_o. In this
case, $l_n(\theta)$ converges towards $E_\theta [\ln f(u,\theta)]$ and θ_n is the ML
estimator converging towards θ_o if $E_\theta [\ln f(u,\theta)]$ has a
unique maximum at θ_o. The last condition is the classical
condition of identifiability in ML estimation.

Even if $g(u) \neq f(u,\theta_o)$, $g(u)$ may be parameterized in such a
way that θ_o and θ_* may be identical in some components. As
a simple example consider a normal linear regression model
with fixed regressor matrices $X_1 \sim nxp_1$ and $X_2 \sim nxp_2$ and
dependent vector $y \sim nx1$.

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon, \quad \varepsilon \sim N(0,\sigma^2 I) \qquad (2.3)$$

The QML estimator β_1 based only on the misspecified model
$y = X_1\beta_1 + \tilde{\varepsilon}$ will converge almost surely to

$$\beta_{1*} = \beta_1 + (X_1^T X_1)^{-1} X_1^T X_2 \beta_2 \qquad (2.4)$$

with $\beta_{1*} = \beta_1$ only if either $\beta_2 = 0$ (no misspecification!) or if $X_1^T X_2 = 0$, that is X_1 and X_2 are orthogonal. Although β_1 will be correctly estimated in the second case, the model will still be misspecified yielding inconsistent estimates not for β_1 but for σ^2 through the heteroscedasticity introduced by $X_2\beta_2$. This example shows why one is interested in tests against general misspecification as well as tests against inconsistency of certain parameters of special relevance.

If almost sure convergence of $\bar{\theta}_n$ towards θ_* and suitable regularity conditions to ensure asymptotic normality are assumed, the asymptotic covariance matrix is derived by a Taylor expansion about θ_*.

$$\frac{\partial l_n(\bar{\theta})}{\partial\theta} = \frac{\partial l_n(\theta_*)}{\partial\theta} + \frac{\partial^2 l_n(\tilde{\theta})}{\partial\theta\partial\theta^T} (\theta-\theta_*) \text{ with} \qquad (2.5)$$

$\bar{\theta}$ QML estimator

$\tilde{\theta} \in L(\bar{\theta},\theta_*)$ with $L(\bar{\theta},\theta_*)$ as the line segment between $\bar{\theta}$ and θ_*.

Since $\partial l_n(\bar{\theta})/\partial\theta = 0$, multiplication by $\sqrt{n}$ yields:

$$\sqrt{n}\ (\bar{\theta}-\theta_*) = -(\frac{\partial^2 l_n(\tilde{\theta})}{\partial\theta\ \partial\theta^T})^{-1}\ \sqrt{n}\ (\frac{\partial l_n(\theta_*)}{\partial\theta}) \qquad (2.6)$$

θ_* minimizes the KLIC $I(g;f(u,\theta))$ or - equivalently - maximizes $E_g(\ln f(u,\theta))$. Hence, one finds for the expected value and the covariance matrix of the score function:

$$E_g(\sqrt{n} \ \frac{\partial l_n(\theta_*)}{\partial\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \frac{\partial \ln f(u,\theta_*)}{\partial\theta} \ g(u)d\lambda(u)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ E_g[\ln f(u,\theta_*)] = 0$$

$$V(\sqrt{n} \ \frac{\partial l_n(\theta_*)}{\partial\theta}) = \frac{1}{n} \sum_{i=1}^{n} \int \frac{\partial \ln f(u,\theta_*)}{\partial\theta} \ \frac{\partial \ln f(u,\theta_*)}{\partial\theta^T} \ g(u)d\lambda(u)$$

$$= \int \frac{\partial \ln f(u,\theta_*)}{\partial\theta} \ \frac{\partial \ln f(u,\theta_*)}{\partial\theta^T} \ g(u)d\lambda(u) \equiv B(\theta_*) \qquad (2.7)$$

Since $\hat\theta_n$ converges to θ_*, the observed information matrix in $\hat\theta \in L(\theta,\theta_*)$ converges to the expected information matrix

$$\frac{\partial^2(l_n(\hat\theta)}{\partial\theta \ \partial\theta^T} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \ln f(u,\hat\theta)}{\partial\theta \ \partial\theta^T}$$

$$\xrightarrow{a.s.} \int \frac{\partial^2 \ln f(u,\theta_*)}{\partial\theta \ \partial\theta^T} \ g(u)d\lambda(u) \equiv A(\theta_*) \qquad (2.8)$$

The main result of this section is the fact that the relation $A(\theta_*) = -B(\theta_*)$ holds true if and only if $g(u) = f(u,\theta_o)$, e.g if the model is correctly specified. This result is obtained immediately by differentiating $\partial \ln f(u,\theta)/\partial\theta$ with respect to θ and computing the expected value.

$$E_g[\frac{\partial^2 \ln f(u,\theta)}{\partial \theta \, \partial \theta^T}]$$

$$= \int [\frac{\partial^2 f(u,\theta)}{\partial \theta \, \partial \theta^T} f(u,\theta) - \frac{\partial f(u,\theta)}{\partial \theta} \frac{\partial f(u,\theta)}{\partial \theta^T}] f(u,\theta)^{-2} g(u) d\lambda(u)$$

The integral

$$\int \frac{\partial^2 f(u,\theta)}{\partial \theta \, \partial \theta^T} f(u,\theta)^{-1} g(u) d\lambda(u)$$

will be 0 only, if $g(u) = f(u,\theta_o)$ with $\theta = \theta_o$.

Hence, in general the vector $\sqrt{n} \, (\bar{\theta}_n - \theta_*)$ will be asymptotically normally distributed with asymptotic covariance matrix $\Sigma(\theta_*) = A(\theta_*)^{-1} B(\theta_*) A(\theta_*)^{-1}$ which is simplified by $A(\theta_*) = -B(\theta_*)$ only if the model is correctly specified. In practice $A(\theta_*)$, $B(\theta_*)$ will be consistently estimated by

$$B_n(\bar{\theta}) = \frac{1}{n} \sum_{i=1}^{n} (\frac{\partial \ln f(u_i,\bar{\theta})}{\partial \theta}) (\frac{\partial \ln f(u_i,\bar{\theta})}{\partial \theta^T} \qquad (2.9)$$

$$A_n(\bar{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \ln f(u_i,\bar{\theta})}{\partial \theta \, \partial \theta^T}$$

The result that $A(\theta_*) = -B(\theta_*)$ only if no misspecification is present is the basis of White's (1982) information matrix test against general misspecification. For the construction of a test against parameter inconsistency the following results about QML estimation with weights are useful.

Assume now that $U_i = (Y,X)_i$ with true density $\tilde{g}(y_i,x_i) = g(y_i|x_i)p(x_i)$ and supposed density $\tilde{f}(y_i,x_i) = f(y_i|x_i,\theta)p(x_i)$ both of which are dominated by the Lebesgue measure. $p(x_i)$ is a non specified non degenerate density. Assume further a positive normed weight function dependent on X only with $\int w(x)p(x)d\lambda(x) = 1$.

Now consider the kernel of the unweighted and the weighted quasi loglikelihood which are maximized with respect to θ.

$$k_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ln f(y_i|x_i,\theta) \qquad (2.11)$$

$$\xrightarrow{\text{a.s.}} \int\int \ln f(y|x,\theta)g(y|x)d\lambda(y)p(x)d\lambda(x)$$

$$\tilde{k}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} w(x_i)\ln f(y_i|x_i,\theta) \qquad (2.12)$$

$$\xrightarrow{\text{a.s.}} \int\int \ln f(y|x,\theta)g(y|x)d\lambda(y)p(x)w(x)d\lambda(x)$$

If the true structure $g(y|x)$ equals $f(y|x,\theta_0)$, then both limits are maximized at θ_0 for any weight function $w(x)$.

However, if the true structure $g(y|x)$ is not a member of the familiy $f(y|x,\theta)$ the limits will be maximized at different points θ_* and θ_{**} unless the functions $p(x)$ and $p(x)w(x)$ are

both members of the subspace of functions that is orthogonal
to the function

$$h(x,\theta_*) = \int \frac{\partial \ln f(y|x,\theta_*)}{\partial \theta} g(y|x)d\lambda(y) \qquad (2.13)$$

with regard to the inner product

$$< h(x,\theta_*),p(x) > = \int h(x,\theta_*)p(x)d\lambda(x).$$

The difference between the unweighted estimator $\bar{\theta}_n$ for θ_* and
the weighted estimator $\overset{\lambda}{\theta}_n$ for θ_{**} forms the basis of the Hausman
type test (Hausman 1978, White 1982).

3. Testing against misspecification

3.1 White's information matrix test

If the model is not misspecified, that is $g(u) = f(u,\theta_0)$ the
matrix $A(\theta_0) = -B(\theta_0)$ or alternatively $A(\theta_0) + B(\theta_0) = O$. To
construct a test of the Wald type, the following notation is
used. The symbol lowtri symbolizes the operation of stacking
the lower triangular of a symmetric matrix in a column vector.
If $A = (a_{ij})$ $i,j=1,\ldots,p$, then lowtri $A = (a_{11},a_{21},a_{22},$
$a_{31},\ldots,a_{pp-1},a_{pp})^T$ will have $q = p(p+1)/2$ elements.

$d(u_i, \bar{\theta}_n) \sim q \times 1 =$

$\text{lowtri } [(\dfrac{\partial \ln f(u_i, \bar{\theta}_n)}{\partial \theta})(\dfrac{\partial \ln f(u_i, \bar{\theta}_n)}{\partial \theta}{}^T) + \dfrac{\partial^2 \ln f(u_i, \bar{\theta}_n)}{\partial \partial \, \partial \theta^T}]$ (3.1)

$D_n(\bar{\theta}_n) = \dfrac{1}{n} \sum\limits_{i=1}^{n} d(u_i, \bar{\theta}_n)$ (3.2)

$D(\theta_*) = \int d(u, \theta_*) g(u) d\lambda(u)$ (3.3)

A Taylor expansion about θ_* yields

$\sqrt{n} \, D_n(\bar{\theta}_n) = \sqrt{n} \, D_n(\theta_*) + \nabla D_n(\hat{\theta})\sqrt{n} \, (\bar{\theta}_n - \theta_*) \text{ with}$ (3.4)

$\nabla D_n(\hat{\theta}) = \dfrac{1}{n} \sum\limits_{i=1}^{n} \dfrac{\partial}{\partial \theta^T} [d(u_i, \hat{\theta})] \sim q \times q \quad \text{ with}$

$\hat{\theta}$ element of the line segment $L(\bar{\theta}_n, \theta_*)$

Using the expansion of equation (2.6) and equation (2.10)
one finds

$\sqrt{n} \, D_n(\bar{\theta}_n) = -[A_n(\tilde{\theta})]^{-1} \dfrac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} \dfrac{\partial \ln f(u_i, \theta_*)}{\partial \theta}$ (3.5)

with $\tilde{\theta}$ element of the line segment $L(\bar{\theta}_n, \theta_*)$.

Combining equation (3.4) and (3.5) gives the result

$$\sqrt{n} \ D_n(\hat{\theta}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{d(u_i,\theta_*) - \nabla D_n(\tilde{\theta}) A_n(\tilde{\theta})^{-1} \frac{\partial \ln f(u_i,\theta_*)}{\partial \theta}\} \ (3.6)$$

With $D_n(\bar{\theta}_n)$ and $A_n(\tilde{\theta})$ converging almost surely against $\underline{D}(\theta_*)$ and $A(\theta_*)$, the asymptotic second moment matrix of $\sqrt{n} \ D_n(\theta_n)$ is given by $M(\theta_*)$:

$$M(\theta_*) = E_g\{[d(u,\theta_*) - \nabla D(\theta_*) A(\theta_*)^{-1} \frac{\partial \ln f(u,\theta_*)}{\partial \theta}] \ \cdot$$

$$[d(u,\theta_*) - \nabla D(\theta_*) A(\theta_*)^{-1} \frac{\partial \ln f(u,\theta_*)}{\partial \theta}]^T\} \qquad (3.7)$$

$M(\theta_*)$ is consistently estimated by

$$M_n(\bar{\theta}_n) = \frac{1}{n} \sum_{i=1}^{n} \{[d(u_i,\bar{\theta}_n) - \nabla D_n(\bar{\theta}_n) A_n(\bar{\theta}_n)^{-1} \frac{\partial \ln f(u_i,\bar{\theta}_n)}{\partial \theta}] \cdot \ (3.8)$$

$$[d(u_i,\bar{\theta}_n) - \nabla D_n(\bar{\theta}_n) A_n(\bar{\theta}_n)^{-1} \frac{\partial \ln f(u_i,\bar{\theta}_n)}{\partial \theta}]^T\}$$

If the model is not misspecified, the expected value of $\sqrt{n} \ D_n(\theta_n) = \sqrt{n} \ D(\theta_0) = 0$ and the matrix $\underline{M}(\theta_0)$ will be the asymptotic covariance matrix of $\sqrt{n} \ D_n(\theta_n)$. The central result of White (1982) is the asymptotic normal distribution of $\sqrt{n} \ D_n(\theta_n)$ and hence the central chi square distribution of the Wald statistic

$$W_n = n D_n(\bar{\theta}_n)^T [M_n(\bar{\theta}_n)]^{-1} D_n(\bar{\theta}_n) \qquad (3.9)$$

with q degrees of freedom, if the model is correctly specified. A necessary condition for the Wald statistic to exist is the non singularity of $M_n(\theta_n)$ which can in practice always be ensured by deleting linear dependent rows and columns in $M_n(\theta_n)$ and the corrresponding rows in $D_n(\theta_n)$ and reducing the degrees in freedom accordingly.

The Wald statistic is very cumbersome to compute because of the third derivatives of the quasi loglikelihood function in the asymptotic covariance matrix $M(\theta_*)$. However, Lancaster (1984) has shown a way to estimate $M(\theta_*)$ without computing third derivatives if the model is corretly specified. Let Z_1 denote the nxq matrix with $z_{i1} = d(u_i,\theta_n)^T$, i=1,...,n and let Z_2 denote the nxp matrix of first derivatives with $z_{i2} = \partial lnf(u_i,\theta_n)/\partial\theta^T$, i=1,...,n. Lancaster's main result is that $M(\theta_*)$ may be estimated by

$$V_n(\theta_n) = n^{-1}(\bar{Z}_1^T\bar{Z}_1 - \bar{Z}_1\bar{Z}_2(\bar{Z}_2^T\bar{Z}_2)^{-1}\bar{Z}_2^T\bar{Z}_1) \tag{3.10}$$

Since $D_n(\theta_n) = n^{-1}\underline{1}^T\bar{Z}_1$ with $\underline{1} \sim$ nx1 is a vector of one's, the statistic of equation (3.9) may be written as

$$W_n = n^{-1} \underline{1}^T\bar{Z}_1[V_n(\theta_n)]^{-1}\bar{Z}_1^T\underline{1} \tag{3.11}$$

$$= \underline{1}^T\bar{Z}_1(\bar{Z}_1^T\bar{Z}_1 - \bar{Z}_1\bar{Z}_2(\bar{Z}_2^T\bar{Z}_2)^{-1}\bar{Z}_2^T\bar{Z}_1)^{-1}\bar{Z}_1^T\underline{1} \tag{3.12}$$

which is easily computed by regressing $\bar{Z}_1$ against $\bar{Z}_2$ and inverting the SSE matrix $nV_n(\theta_n)$. As pointed out above, White's information matrix test is a test against general misspecification. If the hypothesis of no misspecification

is rejected, at least the usual ML covariance estimators $-A_n(\bar{\theta}_n)^{-1}$ or $B(\bar{\theta}_n)^{-1}$ will not be consistent estimators of the covariance matrix of parameters θ_n. Also inconsistency of the QMLE for parameters of interest may be implied. However, as shown before in the context of a simple regression model, model misspecification in some of the parameters of the model does not necessarly result in inconsistent estimation of the parameters of interest. To detect inconsistency of parameter estimates, tests of the Hausman type (Hausman, 1978; White 1982) are used.

3.2 Hausman tests with weight functions

As shown in section 2, the unweighted and the weighted QML estimators $\bar{\theta}_n$ and $\tilde{\theta}_n$ are both consistent estimators of θ_0, if the model is correctly specified. These two estimators are examples of a consistent estimator obtained by ML and by QML methods. However, if the model is misspecified, the ML and the QML estimator will generally diverge. White (1982) refers to any test, that is based on the difference between MLE and QMLE as a Hausman test because Hausman (1978) was the first author to advocate such tests. For the present purpose, I concentrate on Hausman tests that rely only on differences induced by weight functions. Let $\bar{\theta}_n$ and $\tilde{\theta}_n$ be the QML estimates maximizing the quasi loglikelihood functions of equations (2.11) and (2.12). Assuming that the weight function $w(x)$ is non orthogonal in the sense of equation (2.13), the probability limits θ_* of $\bar{\theta}_n$ and θ_{**} of $\tilde{\theta}_n$ will not be equal, unless $g(y|x)$ is an element of the family $f(y|x,\theta)$. Hence the difference $(\tilde{\theta}_n - \bar{\theta}_n)$ is considered which converges almost surely to $\theta_{**} - \theta_*$ which is 0 only if both $\tilde{\theta}_n$ and $\bar{\theta}_n$ are consistent estimators of θ.

The asymptotic distribution of $(\hat{\theta}_n - \bar{\theta}_n)$ is derived by looking at the joint distribution of $(\bar{\theta}_n - \theta_*)$ and $(\hat{\theta}_n - \theta_{**})$. A Taylor expansion about θ_* and θ_{**} in analogy to equations (2.5) and (2.6) yields with $p_i(\theta) = \partial \ln f(y_i | x_i, \theta)/\partial\theta \sim p \times 1$ and $Q_i(\theta) = \partial^2 \ln f(y_i | x_i, \theta)/\partial\theta\partial\theta^T \sim p \times p$:

$$
\begin{bmatrix}
\sqrt{n}\,(\bar{\theta}_n - \theta_*) \\[2ex]
\sqrt{n}\,(\hat{\theta}_n - \theta_{**})
\end{bmatrix}
=
\begin{bmatrix}
-[\frac{1}{n} \sum_i Q_i(\hat{\theta})]^{-1} \frac{1}{\sqrt{n}} \sum_i p_i(\theta_*) \\[3ex]
-[\frac{1}{n} \sum_i Q_i(\tilde{\theta})w(x_i)]^{-1} \frac{1}{\sqrt{n}} p_i(\theta_{**})w(x_i)
\end{bmatrix}
\tag{3.13}
$$

with $\hat{\theta} \in L(\bar{\theta}_n, \theta_*)$ and $\tilde{\theta} \in L(\hat{\theta}_n, \theta_{**})$. If we assume that the following means converge almost surely to their limits given by

$$
A_n(\hat{\theta}) = \frac{1}{n} \sum_i Q_i(\hat{\theta}) \xrightarrow{\text{a.s.}} E_g[Q_i(\theta_*)] = A(\theta_*)
\tag{3.14}
$$

$$
B_n(\theta_*) = \frac{1}{n} \sum_i p_i(\theta_*) p_i(\theta_*)^T \xrightarrow{\text{a.s.}} E_g[p_i(\theta_*) p_i(\theta_*)^T] = B(\theta_*)
\tag{3.15}
$$

$$
G_n(\tilde{\theta}) = \frac{1}{n} \sum_i Q_i(\tilde{\theta})w(x_i) \xrightarrow{\text{a.s.}} E_g[Q_i(\theta_{**})w(x)] = G(\theta_{**})
\tag{3.16}
$$

$$
H_n(\theta) = \frac{1}{n} \sum_i p_i(\theta_{**}) p_i(\theta_{**})^T w(x_i)^2
$$

$$
\xrightarrow{\text{a.s.}} E_g[p_i(\theta_{**}) p_i(\theta_{**})^T w(x_i)^2] = H(\theta_{**})
\tag{3.17}
$$

$$R_n(\overset{\curvearrowright}{\theta}, \overset{\sim}{\theta}) = \frac{1}{n} \sum_i p_i(\overset{\curvearrowright}{\theta}) p_i(\overset{\sim}{\theta})^T w(x_i)$$

$$\xrightarrow{\text{a.s.}} E_g[p_i(\theta_*)p_i(\theta_{**})^T w(x_i)] = R(\theta_*, \theta_{**}) \qquad (3.18)$$

the asymptotic covariance matrix of the vector $\sqrt{n}((\overset{-}{\theta}_n - \theta_*)^T,$
$(\overset{\curvearrowright}{\theta}_n - \theta_*)^T)^T$ is given by

$$(3.19)$$

$$K(\theta_*, \theta_{**}) = \begin{bmatrix} A(\theta_*)^{-1}B(\theta_*)A(\theta_*)^{-1}, & A(\theta_*)^{-1}R(\theta_*, \theta_{**})G(\theta_{**})^{-1} \\ \\ G(\theta_{**})^{-1}R(\theta_*, \theta_{**})^T A(\theta_*)^{-1}, & G(\theta_{**})^{-1}H(\theta_{**})^{-1}G(\theta_{**})^{-1} \end{bmatrix}$$

Hence, the difference $\sqrt{n}(\overset{\curvearrowright}{\theta}_n - \overset{-}{\theta}_n)$ will asymptotically follow a normal distribution with expected value $\sqrt{n}(\theta_{**} - \theta_*)$ and covariance matrix $V(\theta_{**}, \theta_*)$.

$$V(\theta_{**}, \theta_*) = G(\theta_{**})^{-1}H(\theta_{**})G(\theta_{**})^{-1} + A(\theta_*)^{-1}B(\theta_*)A(\theta_*)^{-1} \qquad (3.20)$$

$$-G(\theta_{**})^{-1}R(\theta_*, \theta_{**})^T A(\theta_*)^{-1} - A(\theta_*)^{-1}R(\theta_*, \theta_{**})G(\theta_{**})^{-1}$$

$V(\theta_{**}, \theta_*)$ is estimated consistently by substituting $A_n(\overset{-}{\theta}_n)$, $B_n(\overset{-}{\theta}_n)$, $G_n(\overset{\curvearrowright}{\theta}_n)$, $H_n(\overset{\curvearrowright}{\theta}_n)$ and $R(\overset{-}{\theta}_n, \overset{\curvearrowright}{\theta}_n)$ into their expected values.

As shown by White (1982), the Hausman statistic

$$h_n = n(\tilde{\vartheta}_n - \bar{\theta}_n)^T V(\tilde{\vartheta}_n, \bar{\theta}_n)^{-1}(\tilde{\vartheta}_n - \bar{\theta}_n) \qquad (3.21)$$

will follow a central chi square distribution with p degrees of freedom, if the model is not misspecified in the sense, that $\tilde{\vartheta}_n$ and θ_n estimate consistently θ_o. Since the covariance matrix is combersome to compute from equations (3.15) to (3.18) it may be useful to estimate V in equation (3.20) by applying Lemma 2.1 of Hausman (1978). If the model is correctly specified, e.g. $g(y|x) = f(y|x, \theta_o)$, then $\sqrt{n}(\theta_n - \theta_o) \overset{A}{\sim} N(0, B(\theta_o)^{-1})$ and $\sqrt{n}(\tilde{\vartheta}_n - \theta_o) \overset{A}{\sim} N(0, H(\theta_o)^{-1})$. If the model is not misspecified θ_n will be the ML estimator attaining the asymptotic Rao-Cramér bound for the covariance matrix of $\sqrt{n}(\theta_n - \theta_o)$. The covariance matrix $H(\theta_o)$ will be greater than $B(\theta_o)$ and $H(\theta_o) - B(\theta_o)$ will be positive semidefinite. On the other hand, it may be shown, that the difference $\sqrt{n} \ q_n = \sqrt{n}(\tilde{\vartheta}_n - \bar{\theta}_n)$ is uncorrelated with $\sqrt{n}(\theta_n - \theta_o)$. Hence for the covariance matrix of $\sqrt{n} \ q_n$ which is denoted by $V(\sqrt{n} \ q_n)$ we find with

$$\sqrt{n}(\tilde{\vartheta}_n - \theta_o) = \sqrt{n}(\bar{\theta}_n - \theta_o) + \sqrt{n} \ q_n$$

$$H(\theta_o)^{-1} = B(\theta_o)^{-1} + V(\sqrt{n} \ \tilde{q}_n) \qquad (3.22)$$

Hence, a simple test statistic is given by

$$h_n = n(\tilde{\vartheta}_n - \bar{\theta}_n)^T [H_n(\tilde{\vartheta}_n)^{-1} - B(\theta)^{-1}](\tilde{\vartheta}_n - \bar{\theta}_n) \qquad (3.23)$$

While the matrix in equation (3.20) is always estimated by a
positive definite matrix, the positive definiteness of the
estimated covariance matrix in equation (3.23) depends crucially
on the correct specification of the model. Hence, if the model
is misspecified, empirical researchers will often find the
estimated covariance matrix will not be positive definite.

The power of a Hausman test depends primarily on the weight
function w(x). As Hausman (1978) and White (1982) point out,
the weights should be chosen in such a way that the difference
$(\theta_{**} - \theta_*)$ is large, if misspecification is present. Intuitively
such a behavior is to be expected if the weights are chosen
inversely to the fit of the model. Data points which are fitted
well should be given low weights and points which are fitted
badly should have high weights. The choice of such weights is
discussed in the next section within the context of parametric
rate models.

4. Application to parametric rate models

L $\overset{e}{\cancel{i}}$t T_i, i=1,...,n be continous independent random variables
describing length of time until an event occurs. The density
of the event times is characterized by the transition $r_i(t_i)$
which is parameterized in $\beta \sim$ px1 with explanatory variables x_i.
(Cf. Kalbfleisch and Prentice, 1980 ; Tuma and Hannan, 1984).
Here only the most simple case of a single event is considered
which may readily be generalized to repeated events. The
starting value of T_i is set to 0.

$$r_i(t_i) \qquad = g(t_i, \beta) \geq 0 \qquad \text{parametric rate model} \quad (4.1)$$

$$S_i(t_i) \qquad = \exp(-\int_0^{t_i} r_i(\tau)d\tau) \quad \text{survivor function} \qquad (4.2)$$

$$f(t_i | r_i(t_i)) = r_i(t_i)S_i(t_i) \qquad \text{density function} \qquad (4.3)$$

Maximum likelihood estimators of β are computed by maximizing the loglikelihood function with respect to β by taking into account the effect of right censoring. The individual log-likelihood $l_i(\beta | t_i)$ is given by:

$$l_i(\beta | t_i) = c_i \ln r_i(t_i) - \int_0^{t_i} r_i(\tau)d\tau \qquad \text{with} \qquad (4.4)$$

$$c_i = \begin{cases} 1 & \text{if} \quad t_i \text{ is the time of an event} \\ 0 & \text{if} \quad T_i \text{ is censored at } t_i. \end{cases}$$

To apply the tests discussed in the previous section, one has to compute the first and second derivatives of $l_i(\beta | t_i)$ with respect to β for the unweighted as well as the weighted loglikelihood functions. The weights may be chosen by considering that the deviance increments d_i discussed in the companion paper (Arminger, 1986) are good indicators of the fit of the model for the individual data points. The formula for the deviance increments is given by:

$$d_i = 2\{c_i[\ln(t_i^{-1})-1] - c_i \ln \hat{r}_i(t_i) + \int_0^{t_i^-} r_i(\tau)d\tau \qquad \text{with} \qquad (4.5)$$

$\bar{r}_i(t_i)$ as the ML estimate of $r_i(t_i)$ from the unweighted log-likelihood function.

Since the weights should be functions of the explanatory variables x_i and not of the dependent variable, the deviance increments d_i may either be thought of as prior weights or may be considered as functions of x_i. Using an argument analogous of White (1980a, 1980b) the deviance increments are regressed on a vector of ones, the first derivatives and on their cross-products and the fitted values $\bar{d}_i$ are used as prior weights. Since the weights must be positive, a generalized linear model for d_i with a Gamma distribution as error distribution and the natural logarithm as link funktion is chosen.

As an illustration, I consider a sample of 250 national labor unions in the United states taken from a larger body of data discussed in detail by Freeman et al. (1983) and analyzed previously by Hannan and Tuma (1985). The substantive question is the dependence of the rate of death of a union through disbanding on the size of founding (S measured on a logarithmic scale) and on the duration time. Since new organisations thought to face a higher risk of disbanding than older organizations, a Weibull model with shape parameter $\alpha > 0$ and regression coefficients β_1, β_2 is fitted

$$r(t) = \alpha t^{\alpha-1} \exp(\beta_1 + \beta_2 S) \tag{4.6}$$

The event time is measured in years. The first and second derivatives for such a model are given in the appendix.

Using equation (3.12), a Wald statistic W_n = 165.1 is computed
for White's information matrix test. With three degrees of
freedom this test statistic indicates a strong discrepancy
between the estimates of the information matrix computed
from the first and second derivatives. Assuming now that the
model is misspecified, the covariance matrix $\Sigma(\theta_*)$ derived from
equations (2.7) and (2.8) must be used as covariance matrix for
the estimated parameters.

Table 4.1   Parameter estimates and estimated covariance
matrix from the unweighted loglikelihood function

Estimates	α	β_1	β_2
	0.5742	-1.01746	-0.34031
α	0.00592		symmetric
β_1	-0.02467	0.39115	
β_2	0.00028	-0.04533	0.00720

If the deviance increments d_i obtained from the unweighted QML
estimation are used as prior weights, one finds the following
results.

Table 4.2   Parameter estimates and estimated covariance
             matrix from the weighted loglikelihood function
             with weights d_i

Estimates	α	β_1	β_2
	0.5555	-2.21803	0.04815
α	0.00744		symmetric
β_1	-0.02487	0.11245	
β_2	-0.00049	-0.00017	0.00018

Inspection of the difference between the weighted and the un-
weighted estimators shows only a small difference in α, but
large differences in β_1 and β_2. Computing the Wald statistic for
the Hausman test from equation (3.21) yields a value of 117.3
which is again highly significant with three degrees of freedom.
The information matrix test as well as the Hausman test point
out clearly that the model is very propobly misspecified.
This result is to be expected by using only one explanatory
variable.

Appendix: First and second derivatives of the loglikelihood
of a Weibull rate model

$$r(t) = \alpha t^{\alpha-1} \exp(x\beta), \text{ Weibull rate with} \qquad (A.1)$$

$\alpha \geq 0$ shape parameter

$x \sim 1xp$ row vector of explanatory variables

$\beta \sim px1$ column vector of regression coefficients

Let c_i be the dummy variable indicating right censoring. The individual loglikelihood is then given by:

$$l(\alpha,\beta|t,c) = c\ln\alpha + (\alpha-1)c\ln t + cx\beta - t^{\alpha}\exp(x\beta)$$

$$\frac{\partial l}{\partial \alpha} = \alpha^{-1}c + c\ln t - t^{\alpha}\exp(x\beta)\ln t \sim 1x1$$

$$\frac{\partial l}{\partial \beta} = [c-t^{\alpha}\exp(x\beta)]x^{T} \sim px1$$

$$\frac{\partial^2 l}{\partial \alpha^2} = -\alpha^{-2}c - t^{\alpha}\exp(x\beta)(\ln t)^2 \sim 1x1$$

$$\frac{\partial^2 l}{\partial \alpha \partial \beta^{T}} = -t^{\alpha}\exp(x\beta)x\ln t \sim 1xp$$

$$\frac{\partial^2 l}{\partial \beta \partial \beta^{T}} = -t^{\alpha}\exp(x\beta)x^{T}x \sim pxp$$

Spezifikationsfehler in Regressionsmodellen für Übergangsraten

Hans-Jürgen Andreß

1. Einleitung

Wie diese Tagung in eindrucksvoller Weise demonstriert, werden
Regressionsmodelle für Übergangsraten mittlerweile routinemäßig bei der
Analyse von Ereignis- oder Verlaufsdaten eingesetzt. Die Wahrscheinlichkeit
von Zustandswechseln wird durch einen Set exogener Merkmale und durch
Veränderungen des Prozesses im Zeitablauf prognostiziert. Die neuere
Methodendiskussion dieser Modellklasse untersucht die Konsequenzen einer
Fehlspezifikation, insbesondere die mangelnde Berücksichtigung
unbeobachteter Heterogenität (HECKMAN/ SINGER 1982, TRUSSELL/ RICHARDS
1985, VAUPEL/ YASHIN 1985). Dabei werden alternative Regressionsmodelle und
Schätzverfahren vorgeschlagen, die es erlauben, unbekannte Einflußfaktoren
der Übergangsrate durch beliebige Fehlerverteilungen zu berücksichtigen
(vgl. z.B. FLINN/ HECKMAN 1982, HECKMAN/ SINGER 1984, TUMA 1982).

Leider ist die mathematische Formulierung dieser Modelle so komplex
sowie ihre empirische Überprüfung (Schätzung) so aufwendig, daß sie noch
sehr wenig Verbreitung in der täglichen Forschungspraxis gefunden haben.

Für den Anwender stellt sich insbesonders die Frage, ob die negativen Konsequenzen einer Fehlspezifikation so gravierend sind, daß sich dieser theoretische und empirische Mehraufwand lohnt. In dieser Arbeit wird daher die Frage untersucht, wie stark die Schätzungen eines Regressionsmodells verzerrt werden, wenn ein wesentlicher Bestimmungsfaktor der Übergangsrate unberücksichtigt bleibt.

Nach einigen Definitionen grundlegender Konzepte beschäftige ich mich in Abschnitt 2 zunächst mit der Frage, was unbeobachtete Heterogenität überhaupt ist und welchen Einfluß sie auf die Modellspezifikation hat. In Abschnitt 3 werden dann an Hand eines einfachen Beispiels mit zwei dichotomen Merkmalen und zeitkonstanten Übergangsraten die Konsequenzen der Vernachlässigung des einen Merkmals auf die Schätzungen der Effekte des anderen Merkmals untersucht. An Hand des Modells exponentiell verteilter Wartezeiten lassen sich konkrete Angaben über den Verzerrungsfaktor machen. Sie sind Grundlage der folgenden Monte-Carlo-Simulation in Abschnitt 4, in der an Hand von Stichprobendaten die analytischen Schlußfolgerungen des vorherigen Abschnitts generalisiert werden. Hierbei wird u.a. getestet, ob die Verzerrungen bei Verwendung eines zeitabhängigen (Weibull) bzw. eines partiell parametrischen Modells (Cox) abgemildert oder verstärkt werden. Die Simulation wird mit der Programmiersprache GLIM durchgeführt, da alle drei Modelltypen im Rahmen der Theorie verallgemeinerter linearer Modelle behandelt werden können. Die entsprechenden Programme sind im Anhang dokumentiert.

2. Unbeobachtete Heterogenität

Regressionsmodelle für Verlaufsdaten verwenden als Zielvariable die Rate, mit der Ereignisse im Zeitablauf auftreten. T sei eine positive Zufallsvariable, die die Verweildauer im Ausgangszustand bis zum Eintritt des Ereignisses mißt (Wartezeit). Wenn nur ein bestimmtes (singuläres, nicht-wiederholbares) Ereignis möglich ist, ist diese Rate (Risikofunktion) wie folgt definiert:

$$(1) \quad h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

Sie mißt die Wahrscheinlichkeit, daß im nächsten Moment ein Ereignis auftritt, vorausgesetzt es haben bis dato noch keine Veränderungen stattgefunden. Das Integral dieser Rate bezeichnet man auch als kumulierte Risikofunktion:

$$(2) \quad H(t) = \int_0^t h(\tau) d\tau$$

Weiterhin läßt sich zeigen, daß die Überlebenswahrscheinlichkeit S(t), bis zu einem Zeitpunkt t kein Ereignis zu beobachten, sowie die Dichte f(t), mit der Ereignisse im Zeitablauf auftreten, Funktionen der Rate bzw. der kumulierten Risikofunktion sind:

$$(3) \quad S(t) = \exp[-H(t)]$$

$$(4) \quad f(t) = h(t) \exp[-H(t)]$$

Für einen Prozeß, in dem Ereignisse mit konstanter Rate λ auftreten, ergeben sich dann die bekannten Resultate:

$$(5) \quad f(t) = \lambda \exp(-\lambda t)$$
$$S(t) = \exp(-\lambda t)$$
$$\text{für} \quad h(t) = \lambda$$

Die Wartezeiten T sind exponentiell verteilt und der Anteil der
Untersuchungseinheiten ohne Ereignis ist eine im Zeitablauf fallende
Exponentialfunkion. Der zentrale Parameter beider Verteilungen ist die
zeitkonstante Rate λ. Mit Angaben über die Zeitpunkte von Ereignissen
bzw. über die Zeiten, in denen kein Ereignis aufgetreten ist, kann man
daher die Rate λ durch Maximum-Likelihood-Methoden schätzen.

In einer größeren Stichprobe von Untersuchungseinheiten treten
Ereignisse in den seltensten Fällen mit der gleichen Rate auf. Sie können
im Zeitablauf und mit den Merkmalen der Untersuchungseinheiten variieren.
Es ist daher allgemeine Praxis, die (individuelle) Rate $h_i(t)$ als
(log-lineare) Funktion der Zeit und der exogenen Merkmale aufzufassen:

(6) $\ln h_i(t) = f(\alpha(t), \underline{X}_i(t), \underline{\beta})$

$\alpha(t)$ ist eine zunächst nicht weiter spezifizierte Basisrate, die
Veränderungen des Prozesses im Zeitablauf beschreibt. Je nachdem welche
Restriktionen dieser Funktion auferlegt werden, ergibt sich das partiell
parametrische Modell von COX (1972) oder irgendeine vollständig
parametrische Ratenfunktion (z.B. eine Weibull- oder Gompertz-Rate).
$\underline{X}_i(t)$ ist ein Vektor der exogenen Merkmale der Untersuchungseinheit
i, die wiederum im Zeitablauf variieren können. $\underline{\beta}$ ist der zugehörige
Vektor von Parametern.

Gegen dieses Regressionsmodell ist eingewendet worden, daß es unbekannte
Einflußfaktoren und Meßfehler nur unzureichend berücksichtigt, obwohl es
sich um ein stochastisches Modell handelt (FLINN/ HECKMAN 1982). Es wurde
argumentiert, daß dieses Modell ähnlich wie klassische Regressionsmodelle
mit OLS-Schätzungen einen Fehlerterm enthalten sollte, der diese
unbeobachtete Heterogenität erfasst:

(7) $\ln h_i(t) = f(\alpha(t), \underline{X}_i(t), \underline{\beta}, \varepsilon_i)$

Durch Einführung des Fehlerterms ε_i können die individuellen Raten
$h_i(t)$ bei sonst gleichen Merkmalen $\underline{X}_i(t)$ voneinander abweichen.

Auf der Ebene der gesamten Untersuchungsgruppe ergibt sich die Verteilung der Ereignisse durch Integration der individuellen Dichten:

$$(8) \quad f(t) = \int_{-\infty}^{\infty} f(t|\varepsilon)g(\varepsilon)d\varepsilon$$

$g(\varepsilon)$ ist dabei die Dichte der Fehlerterme. Wie HECKMAN und SINGER (1982) zeigen, können die Schätzungen eines Regressionsmodells für Übergangsraten je nach Verteilungsannahme $g(\varepsilon)$ sehr stark variieren. Da hier verschiedene individuelle Verteilungen zusammengefasst werden, verwendet man auch den Begriff der Mischverteilung. Die Einwände gegen das "deterministische" Modell (6) beziehen sich im wesentlichen darauf, daß der Prozeß auf der Ebene der gesamten Untersuchungsgruppe wesentlich von den individuellen Verläufen abweichen kann. Dieses Problem ist in der folgenden Abbildung 1 veranschaulicht.

Abbildung 1 etwa hier einfügen

Um die Darstellung zu vereinfachen, betrachte ich Regressionsmodelle, in denen nur unbeobachtete Heterogenität auftritt. Obwohl die Raten $h_i(t)$ für alle Untersuchungseinheiten ($i = 1,2,...,N$) verschieden sein können, unterscheide ich lediglich zwei Gruppen von Untersuchungseinheiten, deren Ereignisrisiko $h_1(t)$ bzw. $h_2(t)$ unbekannt ist. $\pi_1(0)$ bzw. $\pi_2(0)$ sei der Anteil dieser beiden Subgruppen an der Gesamtgruppe zu Beginn des Prozesses. Er verändert sich im Laufe der Zeit, je nachdem wieviel Untersuchungseinheiten aus den beiden Subgruppen in ihrem Ausgangszustand überleben (aus Vereinfachungsgründen werden Ereigniswiederholungen nicht berücksichtigt):

$$(9) \quad \pi_1(t) = \pi_1(0)S_1(t)/\left[\pi_1(0)S_1(t)+\pi_2(0)S_2(t)\right]$$
$$\pi_2(t) = \pi_2(0)S_2(t)/\left[\pi_1(0)S_1(t)+\pi_2(0)S_2(t)\right] = 1-\pi_1(t)$$

Das Ereignisrisiko $h(t)$ auf der Ebene der gesamten Untersuchungsgruppe ist dann ein gewichtetes Mittel der subgruppenspezifischen Raten:

$$(10) \quad h(t) = \pi_1(t)h_1(t)+\pi_2(t)h_2(t) = \pi_1(t)h_1(t)+\left[1-\pi_1(t)\right]h_2(t)$$

Abbildung 1 zeigt nun verschiedene Verläufe der subgruppenspezifischen
Raten $h_1(t)$ bzw. $h_2(t)$ und die dazugehörende Rate $h(t)$ der
Gesamtgruppe:

a) Wenn beide Subgruppen eine verschiedene aber zeitkonstante Rate
aufweisen, dann beobachtet man in der gesamten Untersuchungsgruppe eine im
Zeitablauf fallende Rate, weil Untersuchungseinheiten mit hohem Risiko sehr
schnell ausscheiden und die Einheiten zurücklassen, die eine niedrige Rate
aufweisen.

b) Bei vielen sozialen Phänomenen (z.B. Scheidungsraten, berufliche
Mobilität) beobachtet man ein im Zeitablauf zunehmendes Ereignisrisiko, das
jedoch nach einer gewissen Zeit wieder auf einen sehr niedrigen Wert
zurückgeht. Wie Teil b der Abbildung zeigt, läßt sich dieses Verhalten
durch zwei Subgruppen erklären, von denen eine sich praktisch nicht
verändert, während die andere ein im Zeitablauf zunehmendes Risiko aufweist
(Mover-Stayer-Modell). Letztere stellen einen zunehmend geringeren Anteil
der Untersuchungsgruppe, so daß die Rate der Gesamtgruppe nur zu Beginn des
Prozesses von ihnen bestimmt wird und nach einer gewissen Zeit wieder auf
den Wert Null zurückgeht.

c) Aus der Demographie kennt man den J-Verlauf der dritten Rate $h(t)$: Die
Mortalität ist in den ersten Lebensabschnitten besonders hoch
(Säuglingssterblichkeit), verläuft dann mehr oder weniger konstant, um in
den letzten Lebensabschnitten (Altersmorbidität) wieder zuzunehmen. Diese
Situation ist in Teil c dargestellt. Nachdem die Personen der ersten
Subgruppe mit hohem konstanten Risiko ausgeschieden sind, wird der Prozeß
durch die Personen der zweiten Subgruppe bestimmt, deren Rate im Zeitablauf
zunimmt.

d) Summa summarum können die Ergebnisse, die man auf der Ebene der gesamten
Untersuchungsgruppe beobachtet, ein äußerst irreführendes Bild der
tatsächlichen Verhältnisse abgeben, weil man die einzelnen Subgruppen und
ihr Ereignisrisiko nicht kennt. Wie Teil d der Abbildung zeigt, kann

beispielsweise die beobachtete Rate sehr viel langsamer steigen als die subgruppenspezifischen Raten.

Diese und andere Paradoxien werden bei VAUPEL und YASHIN (1985) diskutiert. An dieser Stelle begnügen wir uns mit der Feststellung, daß zeitabhängige Prozesse häufig das Ergebnis unbekannter Risikofaktoren sind. Zeitabhängige Ratenmodelle sind daher weniger Beschreibungen tatsächlicher Veränderungen im Zeitablauf als Approximationen unbekannter Mischverteilungen.

Auch von einem substanzwissenschaftlichen Standpunkt lassen sich in den wenigsten Fällen genuine Zeiteffekte begründen. TUMA und HANNAN (1984) weisen zu Recht darauf hin, daß solche Zeiteffekte in erster Linie das Ergebnis exogener Merkmale sind, die sich im Zeitablauf verändern oder deren Effekte im Zeitablauf variieren. Das gilt auch für solche Variablen, die einen vermeintlich direkten Einfluß der Zeit beschreiben wie z.B. das Merkmal Alter. Alter oder auch Berufserfahrung sind Näherungsgrößen für die unbekannten Faktoren, die sich mit dem Alter oder der Dauer der Berufstätigkeit verändern.

Für den Forschungspraktiker ergibt sich daher die Schlußfolgerung, zunächst möglichst viele Einflußfaktoren zu kontrollieren, bevor zeitabhängige Modelle in Erwägung gezogen werden. Wenn dennoch ein zeitabhängiger Prozeß beobachtet wird, sollte zunächst geprüft werden, ob diese Veränderungen nicht doch das Resultat exogener Merkmale sind, die sich selbst im Zeitablauf verändern. Viel wahrscheinlicher ist jedoch, daß der Modellierungsprozeß schon im ersten Schritt scheitert und wesentliche Einflußfaktoren unberücksichtigt bleiben, weil entweder die entsprechenden Daten fehlen oder weil das theoretische Wissen über den jeweiligen Gegenstandsbereich nur unzureichend ist.

Aus den Erfahrungen bisheriger Forschungspraxis erscheint es daher sinnvoll, zunächst einmal die Konsequenzen einer Fehlspezifikation der einfacheren Modelle zu untersuchen, ehe man in die Diskussion komplexerer zeitabhängiger Modelle einsteigt. Das leitet direkt über zu der Frage, welche Verzerrungen sich bei den im Modell berücksichtigten Merkmalen

ergeben, wenn wichtige Einflußfaktoren außer Acht gelassen werden. Dieses
Problem soll im nächsten Abschnitt diskutiert werden, wobei ich mich
zunächst auf ein zeitkonstantes Modell beschränke. Inwieweit die
Verzerrungen in einem zeitabhängigen Modell verstärkt oder abgemildert
werden, ist Gegenstand des 4. Abschnitts.

3. Spezifikationsfehler in Modellen mit zeitkonstanten Raten

Zur Analyse von Spezifikationsfehlern wird das einfache Beispiel des
vorherigen Abschnitts etwas weiter differenziert: Die beiden Subgruppen J
werden nach einem weiteren dichotomen Merkmal K differenziert, so daß die
gesamte Untersuchungsgruppe in insgesamt 4 Subgruppen zerfällt, die jeweils
mit Häufigkeit n_{jk} auftreten und ein konstantes Ereignisrisiko
λ_{jk} aufweisen (vgl. Tabelle 1). $\pi_{jk}(0)$ seien wiederum die
Anteile der 4 Gruppen zu Beginn des Prozesses, diesmal allerdings bezogen
auf die Gesamthäufigkeit n_j der Subgruppe j=1 bzw. j=2 (Zeilenprozente
der Tabelle).

Tabelle 1 etwa hier einfügen

Alle 4 Subgruppen lassen sich eindeutig durch die beiden Merkmale J und
K beschreiben. Man beachte, daß die beiden Merkmale genau dann unkorreliert
sind, wenn die Anteile $\pi_{11}(0)$ und $\pi_{21}(0)$ identisch sind. Analog
einer Varianzanalyse ergibt sich für die vier Raten folgendes
Regressionsmodell:

$$(11) \quad \ln \lambda_{jk} = \alpha+\beta_j+\gamma_k \qquad \text{mit } \beta_1 = \gamma_1 = 0$$
$$\text{bzw. } \lambda_{jk} = \exp(\alpha+\beta_j+\gamma_k)$$

β_j und γ_k messen den Einfluß der jeweiligen Ausprägung des
Merkmals J bzw. K, wobei der Effekt der ersten Ausprägung von J und K a
priori Null gesetzt wird (Reparametrisierung mit auf eine Kategorie
bezogenen Effekten).

Ähnlich wie im vorherigen Abschnitt für die Gesamtgruppe (vgl. Abbildung
1) soll nun für die beiden Subgruppen J untersucht werden, wie sich ihre
Zusammensetzung ändert, wenn man berücksichtigt, daß beide Subgruppen durch
den Einfluß des Merkmals K in sich heterogen sind. Dazu betrachte ich den
Anteil der jeweils ersten Ausprägung des Merkmals K innerhalb einer
Subgruppe J. Analog Gleichung 9 verändert er sich im Zeitablauf:

(12a) $\pi_{j1}(t) = \pi_{j1}(0)S_{j1}(t)/[\pi_{j1}(0)S_{j1}(t)+\pi_{j2}(0)S_{j2}(t)]$

Unter Verwendung der Überlebenswahrscheinlichkeit (5) eines zeitkonstanten Prozesses ergibt sich weiter:

(12b) $\pi_{j1}(t) = \pi_{j1}(0)\exp(-\lambda_{j1}t)/[\pi_{j1}(0)\exp(-\lambda_{j1}t)+\pi_{j2}\exp(-\lambda_{j2}t)]$

Da sich innerhalb der beiden Subgruppen J die Raten noch einmal nach dem Merkmal K unterscheiden (es sei denn $\gamma_2=0$), verschiebt sich die Zusammensetzung der beiden Subgruppen zugunsten der Ausprägung von K, die die kleinere Rate aufweist. Konsequenterweise wird die subgruppenspezifische Rate $\lambda_{j.}$ in zunehmendem Maße von dieser Unterkategorie k bestimmt. Nach (10) gilt entsprechend:

(13) $\lambda_{j.} = \pi_{j1}(t)\lambda_{j1}+[1-\pi_{j1}(t)]\lambda_{j2}$

Diese Veränderung der Zusammensetzung findet im übrigen nicht mit gleicher Geschwindigkeit in beiden Subgruppen statt (es sei denn $\beta_2=0$). Beide Merkmale J und K können also zu Beginn des Prozesses voneinander unabhängig sein ($\pi_{11}(0) = \pi_{21}(0)$), im Laufe der Zeit werden sie dennoch miteinander korrelieren ($\pi_{11}(t) \neq \pi_{21}(t)$). Dieses Ergebnis ist insofern von Bedeutung, als man aus der Analyse klassischer Regressionsmodelle mit OLS-Schätzungen schließen könnte, daß Spezifikationsfehler bei unkorrelierten exogenen Merkmalen keine Probleme bereiten. Bei Regressionsmodellen mit Raten handelt es sich jedoch um dynamische Modelle, bei denen sich die Zusammensetzung der Untersuchungsgruppe und damit die Zusammenhänge zwischen den exogenen Merkmalen im Zeitablauf verändern. Einige numerische Beispiele sollen diese Entwicklung verdeutlichen.

Tabellen 2 und 3 etwa hier einfügen

In den Tabellen 2 und 3 wird der Verlauf der Anteile $\pi_{11}(t)$ bzw. $\pi_{21}(t)$ sowie der subgruppenspezifischen Raten $\lambda_{1.}$ bzw.

$\lambda_2.$ für ein Regressionsmodell mit den Parametern $\alpha=-1$ und $\beta_2=0.5$ dargestellt. In Tabelle 2 ist die Rate in der Unterkategorie $k=2$ höher als in der Unterkategorie $k=1$ ($\gamma_2=0.5$, positiver Einfluß von K). In Tabelle 3 wird dagegen ein negativer Einfluß von K unterstellt ($\gamma_2=-0.5$). Außerdem wird in beiden Tabellen die Korrelation der beiden Merkmale J und K zu Beginn des Prozesses variiert (unkorrelierte, negativ und positiv korrelierte exogene Merkmale). Bei einem positiven Einfluß von K nimmt der Anteil der ersten Unterkategorie $k=1$ in beiden Subgruppen $j=1$ bzw. $j=2$ zu. Bei einem negativen Einfluß von K ergibt sich genau der umgekehrte Trend. Die Zu- bzw. Abnahme der ersten Unterkategorie vollzieht sich jedoch in unterschiedlichem Tempo, je nachdem welche Subgruppe ($j=1$ oder $j=2$) man betrachtet. Da das Merkmal J in diesem Beispiel einen positiven Einfluß hat, ist die Veränderung in der ersten Subgruppe $j=1$ sehr viel langsamer als in der zweiten $j=2$. In beiden Subgruppen beobachtet man schließlich auf Grund der unberücksichtigten Heterogenität eine im Zeitablauf fallende Rate $\lambda_j.$. Auch hier ändert sich die Rate in der ersten Subgruppe sehr viel langsamer als in der zweiten Subgruppe. Je nachdem ob die beiden Merkmale J und K positiv oder negativ korrelieren, werden die beschriebenen Veränderungen weiter abgeschwächt oder verstärkt.

Für die Analyse von Spezifikationsfehlern ist vor allem die Beobachtung wichtig, daß die Veränderungen in beiden Subgruppen $j=1$ und $j=2$ unterschiedlich verlaufen. Ein einfacher Vergleich dieser beiden Subgruppen ohne Kontrolle des Merkmals K wird daher irreführende Ergebnisse über den Einfluß von J liefern. Dazu betrachte ich das folgende fehlspezifizierte Regressionsmodell, in dem der Einfluß des Merkmals K vernachlässigt wird:

(14) $\ln \lambda_j. = \alpha^* + \beta_j^*$ mit $\beta_1^* = 0$

bzw. $\lambda_j. = \exp(\alpha^* + \beta_j^*)$

Unter Verwendung der Gleichungen 11-13 lassen sich die Parameter α^* und β_2^* des fehlspezifizierten Modells (14) auf die Parameter α, β_2 und γ_2 des wahren Modells (11) zurückführen:

(15) $\lambda_1. = \exp(\alpha^*)$ $= \pi_{11}(t)\lambda_{11} + [1-\pi_{11}(t)]\lambda_{12}$

$= \pi_{11}(t)\exp\alpha + [1-\pi_{11}(t)]\exp(\alpha+\gamma_2)$

$\lambda_2. = \exp(\alpha^*+\beta_2^*) = \pi_{21}(t)\lambda_{21} + [1-\pi_{21}(t)]\lambda_{22}$

$= \pi_{21}(t)\exp(\alpha+\beta_2) + [1-\pi_{21}(t)]\exp(\alpha+\beta_2+\gamma_2)$

Aus Gleichung (15) ergibt sich die relative Verzerrung der Parameter des fehlspezifizierten Modells wie folgt (genauer gesagt: die relative Verzerrung der Antilogarithmen):

(16) $\exp\alpha^* / \exp\alpha =$ $\pi_{11}(t) + [1-\pi_{11}(t)]\exp\gamma_2$

(17) $\exp\beta_2^* / \exp\beta_2 =$ $\{\pi_{21}(t) + [1-\pi_{21}(t)]\exp\gamma_2\} / \{\pi_{11}(t) + [1-\pi_{11}(t)]\exp\gamma_2\}$

Beide Verzerrungsfaktoren sind ebenfalls in den Tabellen 2 und 3 angegeben.

Unabhängig von der Korrelation der exogenen Merkmale zeigt sich, daß der Antilogarithmus der Regressionskonstanten α bei einem positiven Einfluß des vernachlässigten Merkmals K überschätzt und bei einem negativen Einfluß unterschätzt wird. Im Zeitablauf wird dieser Verzerrungsfaktor immer niedriger, so daß bei einem positiven Einfluß von K keine Verzerrungen mehr auftreten, während bei einem negativen Einfluß die langfristige Verzerrung dem Faktor $\exp\gamma_2$ entspricht (vgl. dazu auch Abbildung 2). Anders ausgedrückt, die scheinbare Zeitabhängigkeit des Prozesses auf Grund der unberücksichtigten Heterogenität spiegelt sich in der Regressionskonstanten des fehlspezifizierten Modells wieder.

Abbildung 2 etwa hier einfügen

Die Verzerrung des Effektes β_2 des berücksichtigten Merkmals J hängt von der Entwicklung in beiden Subgruppen j=1 und j=2 ab, wie Gleichung (17) zeigt. Von daher ist die Richtung der Verzerrung schwer abzuschätzen. Man kann sich lediglich überlegen, daß langfristig gesehen keine Verzerrungen auftreten, da Zähler und Nenner in Gleichung (17) für t→∞ entweder den Wert 1 oder den Wert $\exp\gamma_2$ aufweisen. Alle anderen

Werte entnimmt man am besten den Tabellen 2 und 3. Danach wird der Regressionskoeffizient β_2 bei unabhängigen exogenen Merkmalen eher unterschätzt. Bei einem negativen Einfluß von J ($\beta_2 < 0$) ergäbe sich im übrigen eine Überschätzung. Sind die exogenen Merkmale korreliert, dann muß der Einfluß der unberücksichtigten Variablen K bedacht werden:

a) Hat K einen positiven Einfluß und sind beide Merkmale negativ korreliert, dann wird der Regressionskoeffizient β_2 ebenfalls unterschätzt.

b) Hat K einen positiven Einfluß und sind beide Merkmale positiv korreliert, dann wird der Regressionskoeffizient β_2 überschätzt.

c) Wenn K einen negativen Einfluß hat, kehren sich diese beiden Beziehungen gerade um.

Generell gilt, daß die Überschätzung von β_2 im Zeitablauf abnimmt, während die Unterschätzung zunächst zunimmt und dann wieder graduell abnimmt.

4. Simulation von Spezifikationsfehlern - Eine Monte-Carlo-Analyse

Im vorherigen Abschnitt wurde der Erwartungswert der Rate betrachtet.
Regressionsmodelle für Übergangsraten verwenden jedoch Stichproben von
Verlaufsdaten. In diesem Abschnitt soll daher untersucht werden, inwieweit
die Schlußfolgerungen der vorherigen Überlegungen zutreffen, wenn man die
Variabilität des Untersuchungsmaterials berücksichtigt. Darüber hinaus soll
getestet werden, ob der Spezifikationsfehler weniger gravierend ausfällt,
wenn die scheinbare Zeitabhängigkeit des Prozesses (wegen der unzureichend
kontrollierten Heterogenität) durch ein mehr oder weniger zeitabhängiges
Regressionsmodell abgefangen wird.

Zu diesem Zweck wird eine Monte-Carlo-Simulation mit den Parametern des
vorherigen Beispiels (vgl. Tabelle 2 und 3) durchgeführt:

1) Zunächst wird die Stichprobe von insgesamt N Untersuchungseinheiten in 4
Subgruppen der Häufigkeit n_{jk} unterteilt, denen jeweils an Hand des
Regressionsmodells (11) eine subgruppenspezifische Rate λ_{jk}
zugewiesen wird ($\alpha=-1$, $\beta_2=0.5$, $\gamma_2=0.5$ bzw. $\gamma_2=-0.5$).
Die Häufigkeiten n_{jk} werden so gewählt, daß die beiden Merkmale J und K
die gewünschte Korrelation aufweisen.

2) Dann werden mit Hilfe von Gleichung (5) (Überlebenswahrscheinlichkeit) N
exponentiell verteilte Wartezeiten generiert, indem jeweils eine
Zufallszahl u zwischen 0 und 1 gezogen wird, aus der dann an Hand der
subgruppenspezifischen Rate λ_{jk} die Wartezeit t berechnet wird:

(18) $t = \ln u / -\lambda_{jk}$

3) Diese Wartezeiten werden in absteigender Reihenfolge sortiert und die
längsten z Wartezeiten werden zensiert, d.h. ihnen wird eine
Beobachtungsdauer ohne Ereignis zugewiesen, die der Wartezeit der (z+1)-ten
Untersuchungseinheit entspricht (Typ I Zensierung, LAWLESS 1982: 31ff.).

4) Für alle N Untersuchungseinheiten wird eine 0/1-kodierte Dummy-Variable
G gebildet, die zwischen den beiden Subgruppen j=1 (G=0) und j=2 (G=1)
unterscheidet. Der Effekt dieser Dummy-Variablen wird a) in einem
Regressionsmodell mit konstanter Rate (Exponential-Modell), b) in einem
Regressionsmodell mit zeitabhängiger Rate (Weibull-Modell) sowie c) in
einem Regressionsmodell getestet, das lediglich Veränderungen im Zeitablauf
kontrolliert (Cox-Modell):

(19) Exp.-Modell $h(t) = \exp(\alpha + \beta G)$

　　　Weibull-Modell $h(t) = \phi t^{\phi-1} \exp(\alpha + \beta G)$

　　　Cox-Modell $h(t) = \alpha_0(t) \exp \beta G$

Der Effekt β des Exponential-Modells entspricht dem Parameter
β_2^* des o.g. fehlspezifizierten Modells (14).

Diese 4 Schritte werden insgesamt 200 Mal wiederholt und dann werden
Mittelwert, Standardabweichung und mittlerer Fehler (MSQE - mean square
root error) sowie einige weitere Verteilungscharakteristika des Schätzers
β berechnet.

Wie AITKIN/ CLAYTON (1980), WHITEHEAD (1980) und JAGGER/ CLAYTON (o.J.)
zeigen, sind alle drei Modelle (19) Spezialfälle verallgemeinerter linearer
Modelle und können mit der Programmiersprache GLIM (BAKER/ NELDER 1978)
geschätzt werden. Das Exponential- und das Weibull-Modell verwenden die
Statusvariable ST (1=Ereignis/ 0=Zensierung) als poisson-verteilte
Zielvariable mit einem a priori bekannten Offset $\ln(t)$ bzw. $\phi \times \ln(t)$ und
der Dummy-Variablen G als Prädiktor. Der Skalierungsparameter ϕ des
Weibull-Modells wird durch einen iterativen Algorithmus bestimmt. Das
Cox-Modell verwendet die Dummy-Variable G als binomialverteilte
Zielvariable mit einem Offset, der dem logarithmierten Verhältnis der
Risikomengen beider Subgruppen J zu jedem Ereigniszeitpunkt entspricht. Die
Regressionskonstante entspricht hier dem gesuchten Effekt β. Zensierte
Beobachtungen werden bei diesem Modell nicht berücksichtigt und mit 0
gewichtet. Die entsprechenden GLIM-Programme sind im Anhang dokumentiert.

Es wurden insgesamt 54 verschiedene Simulationen durchgeführt, indem der
Stichprobenumfang N (24, 48, 96), der Grad der Zensierung (0%, 25%, 50%),
die Korrelation r der exogenen Merkmale (0, -0.5, 0.5) und der Einfluß
γ_2 des unberücksichtigten Merkmals K (0.5, -0.5) variiert wurde. Die
Ergebnisse zeigen die folgenden Tabellen 4 und 5.

Tabellen 4 und 5 etwa hier einfügen

Auf Grund der Diskussion des vorherigen Abschnitts erwarte ich, daß bei
unkorrelierten exogenen Merkmalen der wahre Einfluß von J in dem
fehlspezifizierten Modell leicht unterschätzt wird, weil die Rate in der
Subgruppe mit höheren Risiko j=2 schneller sinkt als in der
Vergleichsgruppe j=1. Bei einem negativen Effekt von J (z.B. β_2=-0.5)
wäre die Vergleichsgruppe diejenige mit dem höheren Risiko. Folglich müßte
in diesem Fall der wahre Einfluß von J überschätzt werden. Ich habe mich
jedoch in diesem Simulationsexperiment auf einen positiven Effekt von J
(β_2=0.5) beschränkt, d.h. ein vergleichsweise höheres Ereignisrisiko
in der zweiten Subgruppe.

Neben der unterschiedlichen Zusammensetzung der Untersuchungsgruppe im
Zeitablauf spielt aber auch die Korrelation der exogenen Merkmale eine
Rolle: Unabhängig von der Tatsache, daß Veränderungen in den beiden
Subgruppen J mit unterschiedlicher Geschwindigkeit auftreten, ist die
subgruppenspezifische Rate λ_j nach Gleichung (13) ein gewichtetes
Mittel der Raten λ_{j1} und λ_{j2} in den beiden Unterkategorien
jeder Subgruppe. Sind beide Unterkategorien K in beiden Subgruppen J
anteilig gleich häufig vertreten (d.h. J und K unkorreliert) und würden
sich diese Anteile im Zeitablauf nicht verschieben, dann entspräche die
Differenz der beiden logarithmierten Raten $\ln\lambda_2$ und $\ln\lambda_1$.
exakt dem wahren Regressionskoeffizienten β_2. Wenn man nun einmal von
der Verschiebung der Anteile im Zeitablauf absieht, dann läßt sich diese
Differenz auch dadurch verringern oder vergrößern (d.h. β_2 unter-
oder überschätzen), indem man die Unterkategorien K mit dem höheren Risiko
in den beiden Subgruppen J entsprechend erhöht oder verringert. Das ist der
Einfluß der Korrelation der Merkmale J und K, den man in den Tabellen 2 und

3 beobachten konnte:

a) Wenn man die Unterkategorie, die innerhalb von j=2 ein höheres Risiko aufweist, verringert und die Unterkategorie, die innerhalb von j=1 ein höheres Risiko aufweist, erhöht, dann wird die Differenz der beiden logarithmierten Raten kleiner und der Regressionskoeffizient β_2 entsprechend unterschätzt. Dies erreicht man bei einem positiven (negativen) Einfluß von K durch eine negative (positive) Korrelation von J und K.

b) Wenn man die Unterkategorie, die innerhalb von j=2 ein höheres Risiko aufweist, erhöht und die Unterkategorie, die innerhalb von j=1 ein höheres Risiko aufweist, verringert, dann wird die Differenz der beiden logarithmierten Raten größer und der Regressionskoeffizient β_2 entsprechend überschätzt. Dies erreicht man bei einem positiven (negativen) Einfluß von K durch eine positive (negative) Korrelation von J und K.

Wie das numerische Beispiel des vorherigen Abschnitts zeigt, kann der Einfluß der Korrelation der exogenen Merkmale die Verschiebung der Anteile im Zeitablauf überlagern, so daß es zu einer Umkehrung des Trends kommt: Bei positiver (negativer) Korrelation von J und K und positivem (negativen) Einfluß von K ergibt sich im fehlspezifizierten Modell eine Überschätzung des wahren Einflusses von J. Insgesamt läßt sich jedoch feststellen, daß Über- und Unterschätzung des Regressionskoeffizienten β_2 mit zunehmender Zeitdauer immer mehr abnehmen. Anders ausgedrückt, je mehr Untersuchungseinheiten ein Ereignis hatten (je geringer der Zensierungsgrad), desto geringer sind die Verzerrungen von β_2.

Diese Erwartungen werden im wesentlichen durch die Ergebnisse der Monte-Carlo-Simulation bestätigt:

- Mit zunehmender Zensierung nimmt der mittlere Fehler (MSQE) des Schätzers β zu. Die Richtung der Verzerrung bleibt davon aber unberührt.

- Mit zunehmendem Stichprobenumfang nimmt der mittlere Fehler (MSQE) ab.

Der Grund dafür ist weniger eine abnehmende Verzerrung sondern eine abnehmende Streuung der Schätzungen.

- Die Richtung der Verzerrungen entspricht im wesentlichen den Berechnungen des vorherigen Abschnitts. Lediglich bei den unkorrelierten exogenen Variablen zeigt der Mittelwert (Mean) der Schätzer nicht in allen Fällen die erwartete Unterschätzung des wahren Effektes. Eine genauere Analyse der Verteilungsform zeigt jedoch, daß die Verteilung der Schätzer rechtsschief ist, so daß der Median der Verteilung in den meisten Fällen unter 0.5 liegt.Für eine generelle Diskussion von Spezifikationsfehlern ist vor allem bedeutsam, daß die Unterschiede zwischen dem Exponential-, dem Weibull- und dem Cox-Modell nur geringfügig sind. Mit anderen Worten, man ist auch dann nicht vor Spezifikationsfehlern geschützt, wenn man ein zeitabhängiges oder ein partiell parametrisches Regressionsmodell verwendet.

5. Zusammenfassung

In einem Simulationsexperiment mit exponentiell verteilten Wartezeiten und
unterschiedlichen Zensierungsgraden wurde untersucht, welchen Einfluß die
mangelnde Berücksichtigung wichtiger Einflußfaktoren der Rate auf die
Effekte der im Modell verbliebenen Variablen hat. Dabei zeigt sich, daß
Verzerrungen selbst dann auftreten, wenn die vernachlässigten Merkmale von
den berücksichtigten Merkmalen unabhängig sind. Das hängt damit zusammen,
daß sich in einem dynamischen Modell die Zusammensetzung der
Untersuchungsgruppe verschiebt, so daß unkorrelierte Variablen zu Beginn
des Prozesses im Zeitablauf immer mehr miteinander zusammenhängen. Dies
unterscheidet auch Regressionsmodelle mit Übergangsraten von klassischen
Regressionsmodellen, bei denen Spezifikationsfehler dann nicht auftreten,
wenn die exogenen Merkmale unabhängig voneinander sind. Da dies bei
Regressionsmodellen für Übergangsraten offenbar nicht der Fall ist,
verdient die Kontrolle unbeobachteter Heterogenität bei dieser Modellklasse
ganz besondere Aufmerksamkeit.

Natürlich werden auch hier Verzerrungen der Regressionskoeffizienten
verschärft, wenn die unberücksichtigten Merkmale mit den im Modell
verbliebenen korrelieren, was bei sozialwissenschaftlichen Felddaten eher
die Regel sein dürfte. In dem numerischen Beispiel ergaben sich Über- bzw.
Unterschätzungen von bis zu 27%. Diese Verzerrungen blieben auch dann
erhalten, wenn man die scheinbare Zeitabhängigkeit des Prozesses (wegen der
unbeobachteten Heterogenität) kontrollierte oder in einem zeitabhängigen
Modell berücksichtigte. Angesichts des Ausmaßes dieser Spezifikationsfehler
und angesichts mangelnder robuster Alternativen ist es dringend notwendig,
für die Anwender dieser Methoden praktikable Modelle zur Verfügung zu
stellen, die die unbeobachtete Heterogenität der Daten kontrollieren. Das
Modell von TUMA (1982) läßt sich z.B. mit dem Programmpaket RATE (TUMA
1980) umsetzen. Für die Modelle von Heckman u.a. gibt es das allerdings
sehr langsame Programm CTM von HOTZ (1983). In einem weiteren
Simulationsexperiment wäre zu prüfen, inwieweit diese Modelle geringere
Verzerrungen ergeben als die hier betrachteten.

Literaturverzeichnis

Aitkin, M./Clayton, D. (1980): The fitting of exponential, Weibull and extreme value distributions to complex survival data using GLIM. Applied Statistics 29: S. 156-163

Baker, R.J./Nelder, J.A. (1978): The GLIM System Release 3. Generalised linear interactive modelling. Program manual Oxford: Numerical Algorithms Group

Hannan, M.T./Tuma, N.B. (1984): Social dynamics: models and methods. New York: Academic Press

Heckman, J.J./Singer, B. (1982): Population heterogenity in demographic models. In: Multidimensional mathematical demography (Land, K./Rogers, A., eds.). New York: Academic Press. S. 567-599

Heckman, J.J./Singer, B. (1984): A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 52: S. 271-320

Hotz, J. (1983): Continuous time models (CTM): a manual. Pittsburgh, PA: GSIA, Carnegie-Mellon University

Jagger, C./Clayton, D.G. (o.J.): Fitting Cox's regression model to censored survival data. GLIM newsletter

Lawless, J.F. (1982): Statistical models and methods for lifetime data. New York: Wiley

Trussell, J./Richards, T. (1985): Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure. In: Sociological Methodology 1985 (Tuma, N.B., ed.). San Francisco: Jossey Bass, S. 242-276

Tuma, N.B. (1980): Invoking rate. Program manual. Stanford, CA: Stanford University

Tuma, N.B. (1982): Effects of labor market structure on job-shift patterns. In: The analysis of longitudinal labor market data (Heckman, J./Singer, B., eds.). New York: Academic Press

Vaupel, J.W./Yashin, A.I. (1985): The deviant dynamics of death in heterogenous populations. In: Sociological Methodology 1985 (Tuma, N.B., ed.). San Francisco: Jossey-Bass, S. 179-211

Whitehead, J. (1980): Fitting Cox's regression model to survival data using GLIM. Applied Statistics 29: S. 268-275

A N H A N G

Steuerprogramm für eine Simulation

```
$REWIND        1 $INPUT 1 MACROS
$UNITS  24
$DATA   4 NG $READ 6 12 18 24 ! (VIELFACHE VON 4)
$CALC   %P=0.5 : %Z=0
$USE    MAIN
$END
```

GLIM-Macros für die Simulation

```
$SUBFI MACROS
$C ---- MACROS

$MACRO  RATEN ! *** ERZEUGT GRUPPENSPEZ. RATEN ***
$CALC   L= RG(%G)*%LE(NR,NG(%G)) + L*%GT(NR,NG(%G)) : %G=%G-1
$ENDMAC

$MACRO  WEIBULL ! *** SCHAETZT WEIBULL-RATE ***
$CALC   %I=1 : %J=1 : %L=0
$WHILE  %I ITER
$ENDMAC
$MACRO  ITER
$CALC   %K=%J : OFF=%J*LOGT : %L=%L+1
$OFF    OFF $FIT GR
$CALC   %J= (%CU(%YV) / %CU((%FV-%YV) * LOGT) + %K) / 2 :
        %I= %GT((%J-%K) ** 2, 0.0001) : %I=%GT(%L,4)
$ENDMAC

$MACRO  SIMUL ! *** SIMULATION ***

$C      SIMULATE WAITING TIMES

$CALC   T= %LOG(%SR(0)) / -L

$C      SORT DATA

$CALC   GR=G
$SORT   GR GR T : GR GR XNR : T : T T XNR
$CALC   T= ST*T + (1-ST)*T(%Z+1)

$C      CALCULATE OFFSET'S

$CALC   LOGT=%LOG(T)
$CALC   NA=%CU(GR) : NB=%CU(N-GR) : STP=%IF(NA*NB,ST,0) :
        NA=%IF(%GE(NA,1),NA,1) : NB=%IF(%GE(NB,1),NB,1) :
        LOGP=%LOG(NA/NB)
```

```
$C        FIT EXPONENTIAL, WEIBULL AND COX'S MODEL


$YVAR     ST $OFF LOGT $ERR P $FIT GR
          $EXT %PE $CALC E(%S)=%PE(2)
#WEIB     $EXT %PE $CALC W(%S)=%PE(2)
$YVAR     GR $OFF LOGP $WEIGHT STP $ERR B N $FIT $WEIGHT
          $EXT %PE $CALC C(%S)=%PE(1)


$C        ENDE EINER SIMULATION


$OUTPUT 6 $PRINT 'JOB NR. ' %JN 'ITERATION ' %S $OUTPUT
$CALC     %S=%S-1
$ENDMAC


$MACRO    STAT ! *** SIMULATION STATISTICS ***
$SORT     %1
$VARIA    6 %2
$CALC     %2(1)=%CU(%1)/%N : %2(2)= %CU((%1-%2(1))**2) / (%N-1) :
          %2(3)= (%1(Q(1)) + %1(Q(1)+1)) / 2 :
          %2(4)= (%1(Q(2)) + %1(Q(2)+1)) / 2 :
          %2(5)= (%1(Q(3)) + %1(Q(3)+1)) / 2 :
          %2(6)= %SQRT(%CU((%1-0.5)**2)/%N) :
          %1= (%1-%2(1)) / %2(2)
$PLOT     NO %1
$ENDMAC


$MACRO    DESIGN ! *** PRINT IT ***
$PRINT    / :
          'STICHPROBENUMFANG    ' %NU :
          'ANZAHL REPLIKATIONEN' %N :
          'ANZAHL ZENSIERUNGEN ' %Z :
          :
          'GRUPPE  N (KUM.)        RATE'
$LOOK     NG RG
$ENDMAC


$MACRO    MAIN ! *** MAIN PROGRAM ***


$C        INITIALIZE SIMULATION PARAMETERS


$PRINT    'START JOB NR. ' %JN
$VARIA    4 RG
$CALC     RG(1)= %EXP(-1        ) :
          RG(2)= %EXP(-1    +0.5) :
          RG(3)= %EXP(-1+0.5    ) :
          RG(4)= %EXP(-1+0.5+0.5)
$CALC     %N=200 : %S=%N
$DATA     3 Q $READ 50 100 150
$VARIA    %S E W C


$C        GENERATE DESIGN


$CALC     NR=%CU(1) : ST=%GT(NR,%Z) : %G=4 : G=%GT(NR,NG(2)) :
          XNR=-NR : L=1 : N=1
$WHILE    %G RATEN
```

```
$C        DO SIMULATION

$OUTP
$WHILE   %S SIMUL
$OUTP    2

$C RESULTS

$USE     DESIGN
$VARIA   %N NO
$CALC    NO= %ND((3*%CU(1)-1) / (3*%N+1))
$PRINT   : 'STAND. NORMAL PROBABILITY PLOT: EXPONENTIAL MODEL'
$ARGUM   STAT E ES $USE STAT
$PRINT   : 'STAND. NORMAL PROBABILITY PLOT: WEIBULL MODEL'
$ARGUM   STAT W WS $USE STAT
$PRINT   : 'STAND. NORMAL PROBABILITY PLOT: COX MODEL'
$ARGUM   STAT C CS $USE STAT
$PRINT   : '        EXPONENTIAL   WEIBULL    COX ... PARAMETERS'
$LOOK    ES WS CS
$OUTP    6
$ENDMAC
$RETURN
```

Abbildung 1: Unbeobachtete Heterogenität
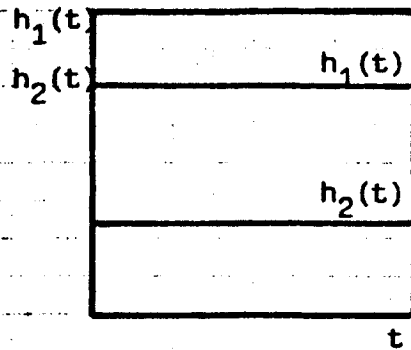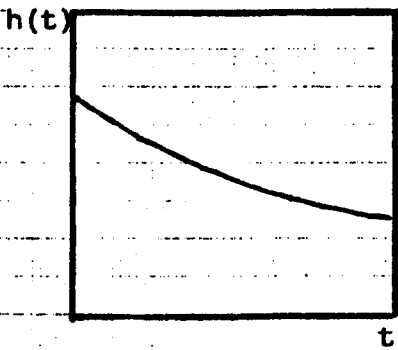
a) $h_1(t)=0.66$, $h_2(t)=0.01$, $\pi_1(0)=0.8$
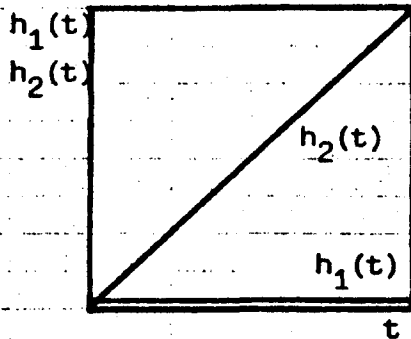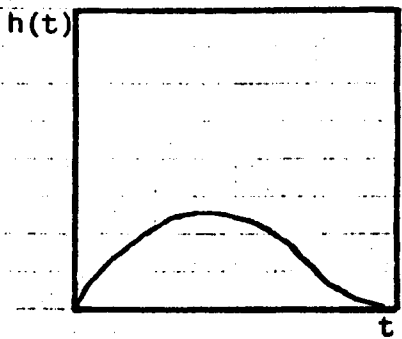
b) $h_1(t)=0$, $h_2(t)=0.002t$, $\pi_1(0)=0.95$
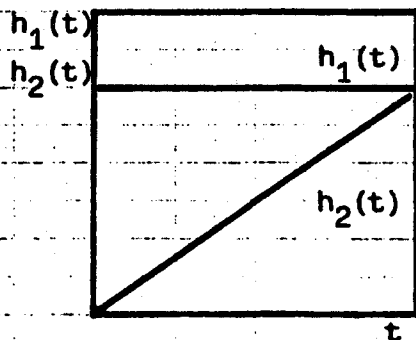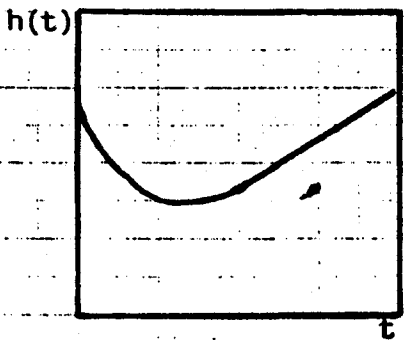
c) $h_1(t)=0.14$, $h_2(t)=0.001+0.0015t$, $\pi_1(0)=0.5$
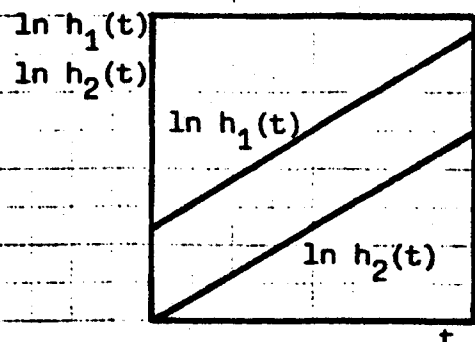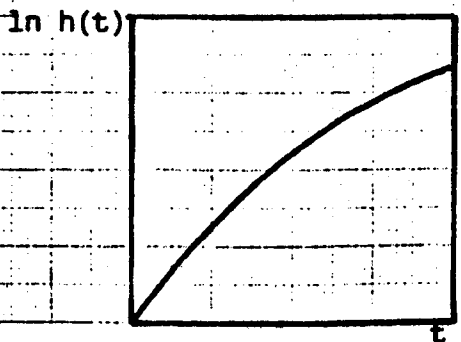
d) $h_1(t)=0.01\exp(0.04t)$, $h_2(t)=0.002\exp(0.04t)$, $\pi_1(0)=0.8$

Tabelle 1: Ein Modell mit zwei dichotomen
Merkmalen und zeitkonstanten Raten

K

	1	2	insgesamt
1	n_{11} $\pi_{11}(0)$ λ_{11}	n_{12} $\pi_{12}(0)$ λ_{12}	$n_{1.}$ 1.00 $\lambda_{1.}$
2	n_{21} $\pi_{21}(0)$ λ_{21}	n_{22} $\pi_{22}(0)$ λ_{22}	$n_{2.}$ 1.00 $\lambda_{2.}$

J

Abbildung 2: Verzerrung der Regressionskonstanten

Fall a: $\gamma_2 > 0$

Fall b: $\gamma_2 < 0$
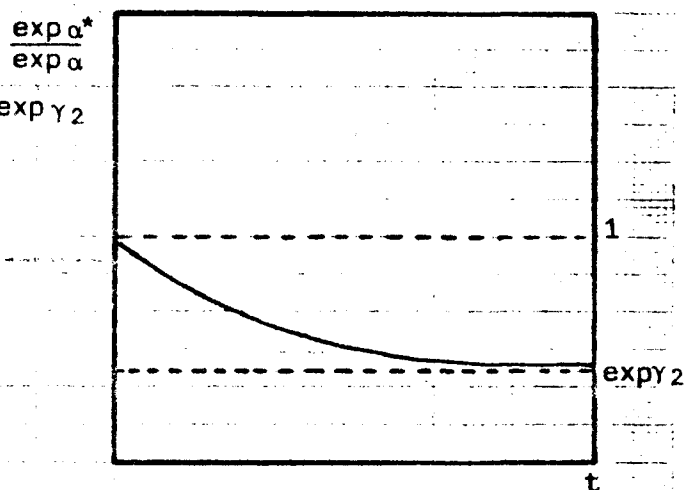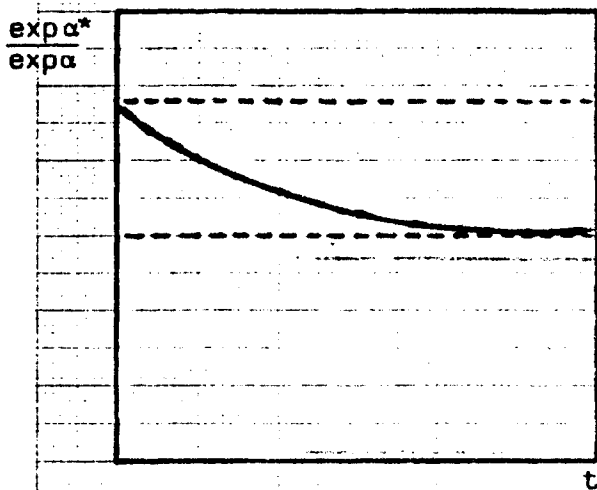
Tabelle 2: Subgruppenspezifische Raten und Verzerrung der Parameter eines fehlspezifizierten Modells ($\alpha=-1$, $\beta_2=0.5$, $\gamma_2=0.5$)

t	π_{11}	π_{21}	$\lambda_{1.}$	$\lambda_{2.}$	α^*	β^*	$\dfrac{\exp\alpha^*}{\exp\alpha}$	$\dfrac{\exp\beta^*}{\exp\beta}$	$S(t)$

unkorrelierte exogene Merkmale r=0.0

t	π_{11}	π_{21}	$\lambda_{1.}$	$\lambda_{2.}$	α^*	β^*	$\dfrac{\exp\alpha^*}{\exp\alpha}$	$\dfrac{\exp\beta^*}{\exp\beta}$	$S(t)$
0.500	0.530	0.549	0.480	0.784	-0.734	0.490	1.305	0.990	0.729
1.000	0.559	0.597	0.473	0.765	-0.749	0.481	1.286	0.981	0.538
2.000	0.617	0.687	0.459	0.730	-0.778	0.463	1.248	0.964	0.302
3.000	0.672	0.765	0.446	0.699	-0.807	0.449	1.213	0.950	0.176
4.000	0.722	0.828	0.434	0.674	-0.834	0.440	1.180	0.942	0.106
5.000	0.767	0.877	0.423	0.655	-0.859	0.436	1.151	0.938	0.066
6.000	0.807	0.914	0.414	0.640	-0.882	0.437	1.125	0.939	0.041
7.000	0.842	0.940	0.406	0.630	-0.902	0.440	1.103	0.942	0.026
8.000	0.871	0.959	0.399	0.623	-0.920	0.446	1.084	0.947	0.017
9.000	0.895	0.972	0.393	0.618	-0.934	0.452	1.068	0.954	0.011
10.000	0.916	0.981	0.388	0.614	-0.947	0.459	1.055	0.960	0.007
20.000	0.992	1.000	0.370	0.607	-0.995	0.495	1.005	0.995	0.000

negativ korrelierte exogene Merkmale r=-0.5

t	π_{11}	π_{21}	$\lambda_{1.}$	$\lambda_{2.}$	α^*	β^*	$\dfrac{\exp\alpha^*}{\exp\alpha}$	$\dfrac{\exp\beta^*}{\exp\beta}$	$S(t)$
0.500	0.273	0.785	0.541	0.691	-0.614	0.244	1.472	0.774	0.734
1.000	0.297	0.816	0.536	0.679	-0.624	0.237	1.456	0.769	0.541
2.000	0.349	0.868	0.523	0.658	-0.648	0.230	1.422	0.763	0.300
3.000	0.405	0.907	0.510	0.643	-0.674	0.232	1.386	0.765	0.169
4.000	0.464	0.935	0.496	0.632	-0.702	0.243	1.348	0.773	0.097
5.000	0.524	0.955	0.482	0.624	-0.731	0.259	1.309	0.786	0.057
6.000	0.583	0.970	0.468	0.619	-0.760	0.280	1.271	0.802	0.034
7.000	0.639	0.979	0.454	0.615	-0.790	0.303	1.234	0.821	0.020
8.000	0.692	0.986	0.441	0.612	-0.818	0.327	1.200	0.841	0.012
9.000	0.741	0.990	0.430	0.610	-0.844	0.351	1.168	0.861	0.008
10.000	0.784	0.994	0.419	0.609	-0.869	0.373	1.140	0.881	0.005
20.000	0.975	1.000	0.374	0.607	-0.984	0.484	1.016	0.984	0.000

positiv korrelierte exogene Merkmale r=0.5

| t	π_{11}	π_{21}	$\lambda_{1.}$	$\lambda_{2.}$	α^*	β^*	$\dfrac{\exp\alpha^*}{\exp\alpha}$	$\dfrac{\exp\beta^*}{\exp\beta}$	$S(t)$
0.500	0.772	0.289	0.422	0.886	-0.862	0.741	1.148	1.273	0.724
1.000	0.792	0.331	0.418	0.870	-0.873	0.734	1.135	1.264	0.534
2.000	0.829	0.423	0.409	0.834	-0.895	0.713	1.111	1.237	0.305
3.000	0.860	0.520	0.401	0.795	-0.913	0.684	1.091	1.202	0.184
4.000	0.886	0.617	0.395	0.757	-0.929	0.651	1.074	1.163	0.115
5.000	0.908	0.704	0.390	0.723	-0.942	0.618	1.060	1.125	0.074
6.000	0.926	0.779	0.385	0.693	-0.953	0.587	1.048	1.091	0.049
7.000	0.941	0.840	0.382	0.670	-0.962	0.561	1.038	1.063	0.032
8.000	0.953	0.886	0.379	0.651	-0.970	0.541	1.031	1.042	0.022
9.000	0.963	0.920	0.377	0.638	-0.976	0.527	1.024	1.027	0.015
10.000	0.970	0.945	0.375	0.628	-0.981	0.516	1.019	1.016	0.010
20.000	0.997	0.999	0.369	0.607	-0.998	0.499	1.002	0.999	0.000

Tabelle 3: Subgruppenspezifische Raten und Verzerrung der Parameter
eines fehlspezifizierten Modells ($\alpha=-1$, $\beta_2=0.5$, $\gamma_2=-0.5$)

t	π_{11}	π_{21}	λ_1.	λ_2.	α^*	β^*	$\dfrac{\exp\alpha^*}{\exp\alpha}$	$\dfrac{\exp\beta^*}{\exp\beta}$	S(t)
				unkorrelierte exogene Merkmale r=0.0					
0.500	0.482	0.470	0.293	0.480	-1.228	0.494	0.796	0.994	0.824
1.000	0.464	0.441	0.290	0.473	-1.237	0.488	0.789	0.988	0.682
2.000	0.428	0.383	0.285	0.459	-1.255	0.477	0.775	0.977	0.474
3.000	0.393	0.328	0.280	0.446	-1.273	0.466	0.761	0.966	0.334
4.000	0.359	0.278	0.275	0.434	-1.291	0.456	0.748	0.957	0.239
5.000	0.327	0.233	0.270	0.423	-1.308	0.448	0.735	0.950	0.173
6.000	0.296	0.193	0.266	0.414	-1.325	0.442	0.723	0.944	0.127
7.000	0.266	0.158	0.262	0.406	-1.341	0.438	0.711	0.940	0.094
8.000	0.239	0.129	0.258	0.399	-1.356	0.436	0.701	0.938	0.070
9.000	0.214	0.105	0.254	0.393	-1.370	0.436	0.691	0.938	0.053
10.000	0.190	0.084	0.251	0.388	-1.384	0.437	0.681	0.939	0.040
20.000	0.052	0.008	0.231	0.370	-1.467	0.472	0.627	0.972	0.003
				negativ korrelierete exogene Merkmale r=-0.5					
0.500	0.237	0.727	0.257	0.541	-1.357	0.744	0.700	1.276	0.820
1.000	0.224	0.703	0.256	0.536	-1.364	0.740	0.695	1.271	0.678
2.000	0.200	0.651	0.252	0.523	-1.378	0.730	0.685	1.259	0.471
3.000	0.178	0.595	0.249	0.510	-1.391	0.717	0.676	1.243	0.336
4.000	0.157	0.536	0.246	0.496	-1.403	0.701	0.668	1.223	0.244
5.000	0.139	0.476	0.243	0.482	-1.414	0.683	0.661	1.201	0.181
6.000	0.123	0.417	0.241	0.468	-1.423	0.663	0.655	1.177	0.136
7.000	0.108	0.361	0.239	0.454	-1.432	0.643	0.649	1.153	0.103
8.000	0.095	0.308	0.237	0.441	-1.440	0.622	0.644	1.130	0.079
9.000	0.083	0.259	0.235	0.430	-1.448	0.603	0.639	1.109	0.061
10.000	0.073	0.216	0.234	0.419	-1.454	0.585	0.635	1.089	0.047
20.000	0.018	0.025	0.226	0.374	-1.488	0.504	0.614	1.004	0.004
				positiv korrelierte exogene Merkmale r=0.5					
0.500	0.736	0.228	0.330	0.422	-1.110	0.248	0.896	0.777	0.828
1.000	0.722	0.208	0.328	0.418	-1.116	0.242	0.891	0.773	0.687
2.000	0.692	0.171	0.323	0.409	-1.129	0.235	0.879	0.767	0.477
3.000	0.660	0.140	0.319	0.401	-1.144	0.230	0.866	0.764	0.333
4.000	0.627	0.114	0.314	0.395	-1.159	0.230	0.853	0.763	0.234
5.000	0.593	0.092	0.309	0.390	-1.175	0.233	0.840	0.765	0.166
6.000	0.557	0.074	0.304	0.385	-1.191	0.238	0.826	0.770	0.119
7.000	0.521	0.059	0.299	0.382	-1.209	0.246	0.812	0.776	0.085
8.000	0.485	0.047	0.293	0.379	-1.226	0.256	0.797	0.784	0.061
9.000	0.449	0.037	0.288	0.377	-1.244	0.268	0.783	0.793	0.045
10.000	0.414	0.030	0.283	0.375	-1.262	0.281	0.769	0.804	0.033
20.000	0.142	0.003	0.244	0.369	-1.412	0.414	0.663	0.917	0.002

Tabelle 4: Mittelwerte und mittlere Fehler in einem fehlspezifizierten
Modell mit exponentiell verteilten Wartezeiten
(positiver Einfluß von K, γ=0.5)

Exogene Variable	Korr. r	N	Zens. %	Exponential-Modell Mean MSQE		Weibull-Modell Mean MSQE		Cox-Modell Mean MSQE		BIAS
	0	24	0	.50	.36	.52	.38	.50	.38	
nicht	0	24	25	.52	.48	.54	.50	.52	.49	positiv
korrel.	0	24	50	.54	.56	.55	.57	.55	.56	
	-.5	24	0	.25	.44	.26	.45	.24	.46	
negativ	-.5	24	25	.28	.51	.30	.53	.29	.52	negativ
korrel.	-.5	24	50	.27	.61	.28	.62	.28	.61	
	.5	24	0	.81	.53	.87	.64	.83	.60	
positiv	.5	24	25	.74	.49	.76	.51	.76	.52	positiv
korrel.	.5	24	50	.77	.68	.79	.70	.77	.69	
	0	48	0	.49	.30	.50	.31	.48	.31	
nicht	0	48	25	.48	.33	.49	.34	.48	.34	negativ
korrel.	0	48	50	.49	.41	.49	.42	.49	.41	
	-.5	48	0	.28	.36	.29	.36	.29	.37	
negativ	-.5	48	25	.24	.42	.24	.43	.24	.42	negativ
korrel.	-.5	48	50	.28	.46	.29	.46	.29	.46	
	.5	48	0	.74	.40	.74	.41	.74	.41	
positiv	.5	48	25	.71	.41	.72	.42	.71	.42	positiv
korrel.	.5	48	50	.76	.52	.76	.53	.76	.53	
	0	96	0	.49	.21	.48	.21	.47	.21	
nicht	0	96	25	.49	.22	.49	.22	.48	.22	negativ
korrel.	0	96	50	.51	.25	.51	.26	.51	.25	
	-.5	96	0	.26	.31	.26	.31	.25	.32	
negativ	-.5	96	25	.23	.35	.23	.35	.23	.35	negativ
korrel.	-.5	96	50	.21	.41	.21	.41	.21	.41	
	.5	96	0	.72	.30	.71	.30	.70	.29	
positiv	.5	96	25	.71	.31	.71	.31	.71	.31	positiv
korrel.	.5	96	50	.77	.41	.78	.41	.77	.41	

Tabelle 5: Mittelwerte und mittlere Fehler in einem fehlspezifizierten Modell mit exponentiell verteilten Wartezeiten (negativer Einfluß von K, $\gamma=-0.5$)

Exogene Variable	Korr. r	N	Zens. %	Exponential-Modell Mean MSQE		Weibull-Modell Mean MSQE		Cox-Modell Mean MSQE		BIAS
	0	24	0	.48	.46	.51	.49	.50	.49	
nicht	0	24	25	.45	.43	.46	.43	.44	.43	negativ
korrel.	0	24	50	.46	.62	.46	.64	.45	.63	
	-.5	24	0	.75	.47	.79	.55	.77	.54	
negativ	-.5	24	25	.74	.53	.77	.57	.76	.57	positiv
korrel.	-.5	24	50	.73	.65	.75	.67	.75	.67	
	.5	24	0	.26	.47	.29	.50	.28	.51	
positiv	.5	24	25	.20	.55	.21	.56	.21	.56	negativ
korrel.	.5	24	50	.26	.61	.27	.61	.27	.61	
	0	48	0	.48	.29	.48	.30	.46	.29	
nicht	0	48	25	.50	.32	.50	.32	.49	.32	(negativ)
korrel.	0	48	50	.50	.41	.51	.42	.50	.42	
	-.5	48	0	.73	.36	.74	.39	.73	.38	
negativ	-.5	48	25	.72	.39	.73	.41	.72	.41	positiv
korrel.	-.5	48	50	.74	.49	.75	.51	.74	.50	
	.5	48	0	.22	.39	.22	.39	.22	.40	
positiv	.5	48	25	.27	.41	.27	.41	.27	.41	negativ
korrel.	.5	48	50	.22	.51	.23	.51	.22	.51	
	0	96	0	.50	.23	.50	.23	.48	.23	
nicht	0	96	25	.51	.24	.51	.24	.51	.24	(negativ)
korrel.	0	96	50	.47	.29	.48	.29	.48	.29	
	-.5	96	0	.74	.32	.73	.32	.72	.31	
negativ	-.5	96	25	.74	.34	.74	.34	.73	.34	positiv
korrel.	-.5	96	50	.77	.39	.76	.40	.77	.40	
	.5	96	0	.23	.34	.23	.34	.22	.34	
positiv	.5	96	25	.24	.34	.24	.34	.24	.34	negativ
korrel.	.5	96	50	.22	.40	.22	.40	.22	.40	

IV Discussions

Discussion of David S. Hachen: "Gender differences in job mobility rates in the United States: A test of human capital and segmentation explanations"

Christof Helberger

David Hachen has presented a very instructive paper on gender differences in mobility rates. It gives a clear picture of various types of voluntary job and status mobility - general mobility, upward authority mobility, leteral and downward authority mobility.

To explain the mobility patterns he uses a set of variables most of which are already familiar from the analysis of income functions to explain earnings and in particular earnings differences between men and women. New, however, is the application of these variables to explain mobility and particularly the arrangement of the explaining variables to the three models he defines to test human capital and segmentation theory.

The result is a useful application of the event history method and reveals insights in various aspects of mobility. In general, the same factors proved to be relevant for the explanation of mobility as of earnings. The main result in terms of gender effects is that the two most prominent theories to explain these gender differences seem to be unsuccessful.

The following comments consist in some minor remarks on
several details of the paper as well as in a more general
consideration. The first remark relates to "part-time work"
as an explanatory variable. I was astonished that the
difference between full-time and part-time work reveals to
be completely irrelevant for upward authority mobility.
This contradicts all my life experience and some empirical
results relating to earnings differences I know from the
literature and from my own research[1].

Hachen points out that women have higher upward mobility in
female dominated occupations compared with male dominated
occupations. This finding seems paradoxical to him and he
tries to explain it by a higher intensity of supervision of
womens' jobs. While this may be an ad-hoc explanation, the
finding is also explainable by the following argument:
assume that the relative frequency of positions with and
without authority (i. e. control intensity) is the same for
occupations dominated by men and by women and assume
further that men are generally privileged in being promoted
to jobs with authority. Then it follows logically that
upward mobility of women in female occupations is higher
than of women in male occupations because there are fewer
men in female occupations which could be chosen for pro-
motion. The chances of women will therefore be higher.

Now to the more general reflections. As an economist I am
primarily concerned with earnings and the earnings
differences between women and men, not so much with job or
status mobility patterns. But I assume that both are
strongly interrelated and that a great part of earnings
mobility of a person is combined with status mobility.
Therefore I mean that status mobility patterns should ge-

---

1) See Mary Corcoran, Greg J. Duncan, Michael Ponza, A
longitudinal analysis of white women's wages, in: J. Human
Resources 18 (1983), S. 510 ff.

nerally be consistent with life cycle earnings patterns.
The stylized facts as regards earnings patterns of women
and men are mainly[2]:

-   the much flatter age-earnings profile of women which
    should correspond to lower upward mobility of women,

-   the disadvantaging effects of interruptions of work
    caused by temporarily withdrawals from the labour
    market,

-   the fact that women without work interruptions have
    nearly the same shape of the age earnings profile - i.
    e. no flatter profile than men -, although they have a
    substantially lower **level** of earnings[3].

The first of these three stylized facts is clearly
confirmed by the lower upward and more frequent sideward
mobility of women which Hachen shows in his paper. The
second fact is not approved by the findings - and I will
present below an attempt to an explanation of this. The
third fact was not subject to test by the models which have
been calculated.

Now, what about the authors'claim of having tested the
explanatory power of human capital and segmentation theory
with respect to gender differences in job mobility? Here I
have some doubts. To begin with the human capital theory.
The approach consists of defining a base line model with
seven explaining variables, one of these is "gender". Then

---

2) See the literature cited by Hachen.

3) See C. Helberger, Humankapital, Berufsbiographie und die
Einkommen von Männern und Frauen, Sonderforschungsbereich
3, Universitäten Frankfurt und Mannheim, Working paper 129
(1984).

the human capital variables "labour market withdrawal" are
added, and it is seen whether the coefficient of gender is
changed. Since for upward mobility the coefficient remains
unchanged it is said that human capital theory cannot ex-
plain gender differences of upward mobility pattern[4].

In this test only a part of the human capital effects is
tested: the effect of withdrawals from the labour market.
Other human capital effects are the influences of
education, labour market experience and duration on the
job[5]. These effects are included in the base line model and
are not tested. A more comprehensive test of the human
capital theory could therefore define a base line model
which includes no human capital variables, then the human
capital variables could be included and it could be seen
what their effect is.

Furthermore, assume that the temporarly withdrawal of a
woman leads to a re-entry into the labour market at a level
which is lower than the level she had before the
interruption. This is an assumption which seems plausible -
at least for a portion of women - and is consistent with
the findings of income regressions of the human capital
model for the USA and other countries (obsolescence of
human capital, loss of previous work experience). Assume
further that there is a restauration effect of human
capital after the interruption as was found in the income
regressions. It is possible that this restauration effect
leads likewise to upward mobility. It could then be that
persons with and without work interruptions have about the

4) The author does not say that human capital could not
explain upward mobility. The thesis is that it cannot
explain gender differences of upward mobility.

5) It is admitted that these variables can also be
interpreted in other ways, e. g. in terms of seniority and
adscriptive traits, but this must not be incompatible with
their human capital interpretation.

same amount of upward mobility, and this could be the cause
why the coefficient of the withdrawal variables in model II
are not significant. Nevertheless there would be
differences in the mobility patterns of persons with and
without withdrawals in that the first end up with lower
positions. This would be concealed by the used measure of
mobility, since it is a measure of **relative** position
changes and not of absolute changes like the income
measure.

Let me now look at the test of the segmentation theory. The
result of model III is that gender differences in upward
mobility are not affected by the fact that a person works
in a predominantly female or a predominantly male
occupation. The distinction between male and female
occupations is interpreted as the influence of segmen-
tation. But assume that men are strongly over-represented
in higher and more authoritative positions and that they
resist to the promotion of women which would lead to upward
mobility of women as is consistent with the findings of
models I to III. This could also be named an effect of
segmentation. A convinced adherent of segmentation theory
could therefore say only a specific effect or variant of
segmentation was tested but that there are other forms of
segmentation which are consistent with or even approved by
the findings. The difficulty is, however - and I agree with
the author in this respect -, that the meaning of the word
"segmentation theory" is not very clear, that there are
many different types of such theories and that they are
often vaguely formulated so that they are difficult to
test.

Discussion of Andreas Diekmann: "Effects of education, occupational characteristics and cohort on the 'family cycle'"

Annemette Sørensen

I was very excited when I first read this paper because Diekmann has an excellent dicussion of the non-monotonic risk distributions that characterize both the marriage and the divorce transition and because he tries to show the link between one of the old models of the marriage transition- developed in the early 1970's by Gudmund Hernes- and the survival models that we use today. Now that I have finished reading the paper, I am still excited about this attempt at translation between models, although I have some questions about the specific formulation of the translation. I am disappointed though by the multivariate analysis which does not seem to reflect the discussion of models in the first part of the paper. Here it is shown that the log-logistic and the Hernes model both provide very good fits to the data on the transition to first marriage; they both reflect the curvilinear age dependence of marriage rates. Nonetheless, in the multivariate analysis, a Cox model is estimated, thus leaving us without an estimate of the duration parameter. I am not arguing that the partial likelihood model gives us biased estimates,- I do not think that is the case; I would like, however, to see an estimate of the duration effect which after all does seem to be of some interest in this context. In particular, I think we need

a discussion of how we should expect the duration effect to change with the introduction of covariates in the model. Should we, for example, expect the age effect to disappear in a perfectly specified model? or should we expect it to increase ? In the former case we expect heterogeneity to be the explanation of the observed duration effect, that is we think we can explain it with variables other than age or duration at risk. The latter might be the case if there are strong preferences for making the transition at a particular age; the covariates in this case then control for variations in peoples' ability to realize this preference. I might add that in the models I have estimated on Norwegian data, it seems to be the case that the duration effect in a log-logistic model increases in the better specified models.

The choice of covariates in the estimated models is largely dictated by expectations derived from economic theory about marriage and divorce. The strongest predictions are about gender differences: educational attainment, occupational prestige and income should be positively related to the risk of marriage and negatively to the risk of divorce for men; for women the opposite is supposed to be the case. These predictions are predicated on two very strong assumptions, namely that the traditional sexual division of labor is compatible with men's desire to marry 'suitable' women, and that the gain from marriage largely is determined by a task specialization where women specialize in domestic work, men in market work. These assumptions might indeed hold for large parts of the population, especially for older cohorts. We might in other words expect a change over time in the extent to which the economic hypotheses are supported by the empirical evidence. It is also likely that such changes over time are more pronounced for some groups than others; the highly

educated men and women in the youngest cohorts very likely do not fit this pattern very well.

This might indeed be part of the reason that many of the predictions do not find empirical support. Another reason,- and Diekman knows this - has to do with his data. The hypothesized gender differences in the effects of educational attainment on the marriage rate are generally supported. The difference is not large, but it is there. I have one question/suggestion to make here: is it possible that the education effect for highly educated young men is increasing over birth cohorts ? This should be the case because even if men are the ones to benefit from the 'ideal economic' marriage, there also is a strong tendency towards homogamy. Men with higher education should want to marry suitable women; in the younger cohorts this is more likely to mean women with higher education (because gender differentials in educational attainment are shrinking); but in younger cohorts these are the women who typically would not buy into the traditional sexual division of labor. The transition for these men should thus be delayed and this should show up as a negative effect of educational attainment for men in younger cohorts.

The test of the hypothesis that occupational prestige and income delays marriage for women but not for men is problematic. These variables are measured at the time of interview,- which for many must be years after they married. As Diekman acknowledges this is problematic in particular for women because the marriage transition itself to a large extent determines women's occupational and economic position. The fact that income for example has a negative effect on the rate of marriage for women and a positive one for men, probably reflects the fact that marriage tends to decrease women's and increase men's income ; it

does not necessarily say much about the effects of income on the marriage decision itself. I find it difficult to see that this provides a test of the economic hypothesis, at least with respect to the gender differences. Unfortunately, the data do not seem to permit such a test.

With respect to the divorce transition, the expected interaction effect between gender and education is found. Highly educated women have higher risks of divorce. Again I would like to see a test of my hypothesis that for men in the younger cohorts, we also see this effect, because they likely are married to those highly educated women.

The last of the substantive concerns in the paper is to estimate how much the increase in educational attainment has contributed to a change in the marriage and divorce transitions. With respect to marriage, the results suggest that the post second World War expansion in education has added about .8 years to the median marriage age for women and only .2 years to that of men's. This result is consistent with economic theory which predicts that educational expansion delays marriage for both women and men because people spend longer attending school. Educational attainment in addition delays marriage further for women, because it decreases their gain from marriage.

I think there is a puzzle here that needs to be addressed. In Table 1 we see the median age at marriage for 5 birth cohorts. For both men and women the median age decreases over the four oldest cohorts and then it increases again for the youngest. For each successive cohort educational attainment is higher- at least I believe that to be the case; we should thus expect the median age at marriage to increase as well. That is, if the economic theory is correct. That is clearly not the case. How do we

explain this ? are there other factors characteristic of these birth cohorts which account for this ? The analysis shown in Table 7 and 8 is somewhat misleading, because it does not alert us to the fact that there also has been a decline in age at marriage during the time period described, --a decline which clearly provides counterevidence to the hypothesis.

Let me close with a few comments and questions about the translation of the Hernes model into transition rate model language. I was as I said at the outset very pleased to see this done. I was also pleased to see the comparison between the fit of the Hernes model and the log-logistic model, and the comparison between the components of the two models. I have always liked the Hernes model because the theoretical justification for it was so much better than any other I know of. He simply says that change in an individual's rate of marriage is a function of the social pressure to marry and the person's marriage potential. Social pressure is assumed to increase with age, or rather with the proportion of a cohort who has already married; marriage potential is assumed to decrease with age. These simple assumptions lead to a model which according to Diekmann easily can be translated into a rate model with the hazard being equal to the product of the proportion already married at time t and the marriage potential at time t. We are assured that this is in fact a non-monotonic hazard rate.

In contrast to the Hernes model, the hazard rate in the log-logistic model is the product of the proportion already married at time t and a ratio p/t,- clearly a factor that is declining with time. It is not clear to me what this p refers to ? what is

the difference in substantive terms between the specification of the hazard rate in the Hernes model and the log-logistic ? It seems that we need to know this in order to make meaningful choices between the two models.

Discussion of Michael Wagner: "Education and migration"

Heiner Meulemann

What are the relations between education and regional mobility? Michael Wagner splits this general question in two specific analyses: How affects <u>regional origin</u> educationally induced <u>mobility</u>? And: What effects has <u>education</u> on <u>migration</u>. These questions are treated using life histories of three German cohorts born around 1930, 1940 and 1950. The answer to the first question is, contrary to expectation, that rural origin lowers educationally induced mobility. The answer to the second question is, in accord with expectation, that higher education furthers regional mobility indepently of <u>stage in the life cycle</u>, which itself shows substantive effects, and of <u>cohort membership</u>, which itself shows somewhat smaller effects; this results hold even if distances of regional mobility are controlled. In a final step of analysis therefore, the effect of education on regional mobility is examined in Cox regression models controlling occupational position, local attachment and life cycle; dependent variable is the rate of mobility, units of analysis are periods of permanent residence until the next move, instead of individuals. In every model there are significant positive effects of education on migration, but these effects are substantively considerable only for long distance migration and for the highest educational group. Out of the controls chosen, only variables of the life cycle reduce the effect of education on migration, although significant and substantive effects of education remain. In sum: education exerts a strong positive influence on migration, partly mediated by, partly competing with effects of

the life cycle.

In my opinion, it is not at all an accident that position in
the life cycle shows up as the only serious competitor to
education in determinating migration. In the context of a
migration study education itself becomes a life cycle varia-
ble; it is the first education which determines moves, not
concurrent education and on the job training. Michael Wagner
has treated the relation between education and migration
with great skill empirically and produced results which give
important clues theoretically. My suggestion, therefore, for
following up the questions he has risen is to focus more
strongly on the life cycle in the theoretical and empirical
analysis of education and migration instead of treating this
relation theoretically in a predominantly static vein, there-
by at the same time using the full potential of the data at
hand. I want to substantiate this in a couple of more de-
tailed suggestions.

(1) The most interesting question treated in the Cox re-
gressions to me seems to be if the specific position in the
life cycle or the attachment to a locality working indis-
criminately over the life cycle affects migration. To this
question the differentiation between background and occupa-
tional variables (between model 1 and 2) is secondary, but
a model which - over and beyond these two groups of variab-
les - controls for attachment or life cycle only is neces-
sary; unfortunately the latter has not been computed and
the former (model 3) is not presented in table 10. From the
results presented in table 10, model 4F, attachment (BELEG,
EIG and OGEB; GEBORT to me should be excluded already for
theoretical reasons) as well as being married (EHE), hav-
ing school age children (SFZ3) and having a marriage par-
tner in the labor force (ERWEHEP) strongly depress migra-
tion rates. How are these effects related? Which are the

more important ones?

(2) As far as I can see, migrations are analyzed indiscriminately although the inventory (page 13) allows a discrimination between parents and own, forced and autonomous, passive and active migrations. But: How far is my parents` migration my own migration? This difference is particularly important in an analysis of the effects of my education on migration? Are migrations when entering school (table 4) moves to attend a particular school or decisions of parents coinciding with school entry by chance? Most probably, the higher the education the later the first autonomous migration takes place. In that case, the dependent may be contaminated with the independent variable. However this may be, it seems worthwile to analyze, instead of any migration, the first autonomous migration; this would have the additional advantage of reducing the sample from spells to individuals and bringing the analysis process closer back to the real life process. The move out of the parents` house is a highly important event in the life cycle and it may have determinants quite different from migrations in later life. If so, it seems altogether questionable to analyze migrations indiscriminately.

(3) It would be good to have a theory of the relations between life cycle and migration, termination of first education being taken as one of the significant events in a normal life career. How is marriage and child birth, graduation and job entry related to migration? What determines migration processes once a marriage partner and a job is found? To analyze these questions it seems worthwile to focus on specific migration processes. One had to divide each individual career in segments: from birth to moving out of the parents house, from moving out of the parents house to marriage, from marriage to first migration and so

forth; or: from birth to marriage, from birth to job entry; these spells need not be mutually exclusive. For each spell, the unit of analysis would be the individuals. In each spell one could choose different determinants as predicators of the migration. Specifically, one could, instead of restricting predictors to the beginning of a spell, construct predictors referring to the life career immediately preceding the event. Has there been a graduation or a marriage in the last, say, fife months before migration? Being recently married might affect migration quite differently from being married. In this manner, one might try to study migration in relation to changes in occupational and marital status, rather than migration per se.

(4) One of the most intriguing of Michael Wagners analyses treats the synchronization between educational events and migration (table 4). In taking the first education as a variable of the life course it might be interesting to analyze the synchronization between other life course decisions, say, marriage and migration, and to compare the two forms of synchronization. Furthermore it might be interesting to analyze the triple synchronization between migration, education and marriage. The results of table 4 show that synchronization between leaving school and migration is decreasing in the younger cohorts. Assuming that this migration is in most cases the first autonomous migration leaving the parents house, one might ask if there is a general trend of less synchronization, less coherence in younger cohorts. For, migrations in general are increasing in younger cohorts - as table 3 shows. Is increasing mobility coupled with decreasing synchronization? Are life careers becoming at the same time more volatile and less coherent? Is decreasing stadardization of life careers a reaction of increasing demands or strains in life careers?

Discussion of Nazli Baydar: "What can backward recurrence time data tell us: An application to residential mobility in the U.S."

Douglas A. Wolf

This is an interesting and thought-provoking paper, one of a handful to appear so far dealing with the issues of estimating failure-time distributions from backwards recurrence-time data. Backwards recurrence-time data, in its most restrictive form, is severely truncated information on a dynamic process: at some fixed instant a survey is taken (or the process is somehow sampled) and for each sample point the elapsed time since the event coinciding with entry into the current state is determined. Since the next event, and its timing, are unknown, each observation represents a right-censored failure-time (i.e. a period of survivorship).

Baydar mentions a number of previous papers in demography dealing with applications of the backwards recurrence-time estimation problem. It should also be noted that for many years labor economists have dealt with essentially the same problem: estimating the distribution of durations of unemployment spells, given data from a cross-sectional sample in which those currently unemployed are asked the number of weeks to date in their present spell of unemployment. Highlights of this literature include the initial work of Kaitz (1970), and the often-cited paper by Salant (1977); see also Nickell (1979) and Flinn (1986). The literature just cited, in combination with the literature cited by Baydar, includes a variety of approaches to the problem; Baydar's contribution adds further to this variety. It seems likely that the question of formulating models for recurrence-time data, still in a state of relative infancy, will undergo considerable further development in the future.

Baydar begins by reviewing the simplest type of renewal process, the Poisson process. The process operates on a time dimension defined by an individual's life-time. Thus the origin of the process is the individual's "birth"--which, in certain applications, may take place at a positive chronological age. From the time of birth, events occur at a constant rate. At calendar time t^* a survey is taken, and for each respondent we determine u--the elapsed time since entry into the current state--and d--a dummy variable equalling 1 if the respondent experienced an event at time $t^* - u$. d will equal 0 if the individual has experienced no events since birth. Baydar presents, as the individual's contribution to the likelihood of the data, the expression

$$L = a^d \exp(-au)S_c(x) \tag{7}$$

where x is the individual's current age, $S_c(x)$ is the probability of surviving one or more censoring mechanisms--in practice, mortality--and a is the unknown rate

parameter of the Poisson process (I use Baydar's equation numbers, but have changed her notation slightly). This expression can be factored into three components: first, a represents the renewal density for the event of interest, which is age-invariant, in the Poisson process; second, $\exp(-au)$ is the survivor distribution associated with the time from the nth to the $n+1$th renewal; and third, $S_c(x)$ is the probability of being alive, and of age x, at the survey date. The product of the first two of these three terms is the density function for backwards recurrence times for a process not subject to censoring.

Baydar goes on to consider more complex models, specifically renewal processes with other than a constant hazard. In this case, the distribution of backwards recurrence times is more complicated. The renewal density—defined, we must recall, on the individual's age—is, in general, nonconstant, although it converges to a constant after "sufficient" time (age) has passed. The value of the renewal density at a given age can be found by inversion of a Laplace transform, itself expressed in terms of the Laplace transform of the failure-time distribution. The individual's contribution to the sample likelihood now becomes

$$L = m(x - u)^d S(u) S_c(x) \qquad (13)$$

where $m(z)$ is the value of the renewal density at age z.

The difficulty with this more general model, Baydar reminds us, is the difficulty of inverting the Laplace transform of the renewal function (differentiation of which yields the renewal density). The difficulty can be avoided if we can reasonably assume that sufficient time has elapsed such that the renewal density has converged (to the inverse of the mean of the failure-time variable). If so, (13) would reduce to

$$L = S(u) S_c(x) \mu^{-d} \quad ,$$

where μ is the mean of the failure-time variable.

Baydar goes on to show how proportional-hazards effects of covariates can be introduced into the model. The models are then applied, using a sample of observations on time-since-last-residential-move from the 1980 U.S. Census. First, results for univariate models are presented, in the form of a series of two-subsample models. The assumed failure-time density in these models is a mixture, the form of which is

$$f(t) = \delta_1 a^2 t \exp(-at) + \delta_2 b \exp(-bt) \quad . \qquad (18)$$

In this series of univariate models, the explicit renewal-density specification of the likelihood function is evidently used [i.e. equation (13)] which, in turn, implies the need to numerically invert the Laplace transform; the specific algorithm used in the numerical inversion is not specified (these are details which need clarification).

The second set of results pertain to a Poisson process with proportional-effects covariates; estimation of this model is straightforward in view of the constancy of the renewal density.

The third set of results pertain to the more general renewal process generated by the failure-time density given in (18); in order to estimate the model, it is assumed that for each sample observation the renewal density has converged to the inverse of the mean failure time. It is interesting to note that the estimated covariate effects in models 2 and 3 are quite similar; parameters found to be significant in one specification have approximately the same magnitude, and standard error, in the other specification.

Baydar's paper clearly represents a large effort, particularly in the area of mathematical manipulation and computation. My comments for the most part pertain to issues of model specification rather than to the substantive application, migration.

A significant proportion of the paper is devoted to the problems associated with evaluating recurrence-time distributions for processes with other than constant hazards. Since processes with constant hazards are rarely thought to be sufficiently general, it is appropriate to focus on the more general case. Yet most of the difficulties are due to the presence of the renewal density in the distribution function for backwards recurrence times. It would seem useful to consider ways to eliminate the renewal density from the problem, and thereby to eliminate the need to invert Laplace transforms.

Note that although the failure-time distribution is age invariant, Baydar conditions her likelihood on an individual's age at the time of the survey. This is because, from the perspective of individual lifetimes, the time dimension for which is age, the renewal density is nonconstant, at least until it can reasonably be assumed to have converged. Note also that, since the sample is treated as observations from iid renewal processes which are age-invariant, we must implicitly assume that the process has been stable for at least several years.

Suppose, instead, that we assume a somewhat different form of stability, namely stability of the population, and of its associated life-cycle dynamic processes, from which the data are drawn. In other words, suppose that from year to year the age structure of the population can be represented by a (calendar) time-invariant density function $g(x)$. Suppose, as before, that events are generated by a renewal process determined by an age-invariant failure-time density $f(t)$; associated with the renewal process is a renewal density $m(x)$--x, as before, represents age. Then at any instant the proportion of the population experiencing an event is

$$\int_0^\infty g(x)m(x)dx \quad ,$$

which is a (calendar) time-invariant constant, r. From the perspective of a fixed survey date, t^*, the probability that a randomly-selected individual, **unconditional** on current age, experienced an event u time units ago is $1/r$. The probability of experiencing an event u units ago, and surviving further events until t^*, is $r^{-1}S(u)$; integrating over the domain of u, the density of backwards recurrence times is $r^{-1}S(u)/r^{-1}\mu$

$$r^{-1}S(u)/r^{-1}\mu$$

$$= S(u)/\mu \quad , \tag{i}$$

where μ is the mean of t. The assumption that the population generating the observations is (and has for some time been) stable allows us to eliminate the renewal density from the expression for the recurrence-time distribution, even though the renewal processes of individuals within the population are not in "equilibrium" [in the sense used by Cox (1972)--i.e., where the renewal density has converged].

Equation (i) in fact serves as the basis for estimation of unemployment durations in several papers [eg. Flinn (1986)]. In such papers, the assumption used to justify the model is that the process is in a "steady state"--the counterpart, apparently, to the stability assumption I have used above.

An open question is whether the population-stability assumption is more or less innocuous than those used by Baydar. I suspect that overall, the two approaches are not very different. However, I have proposed a method in which the age of the respondent at the time of the survey (if available) is not used in the analysis; the two approaches might therefore give different results, and this is a

question worthy of exploration.

A second comment relating to model specification concerns the difference between what Cox calls "ordinary" and "modified" renewal processes; I prefer Ciniar's term, "delayed", for the latter. A delayed renewal process for events of some type, j, is one for which the time origin is the first j-event; or, it is one in which the failure-time distribution for the first renewal differs from that for all subsequent renewals. Baydar has adopted an ordinary renewal model; the time origin is an individual's 18th birthday, and renewals take place at subsequent residential moves. The corresponding delayed renewal process is the one whose time origin is the first residential move after reaching age 18. In many applications the delayed renewal model is a more appropriate specification; in some applications, it is the only appropriate one. For example, the renewal process underlying backwards recurrence time data on interrupted unemployment spells is one whose time origin is the first time an individual becomes unemployed; that event is necessarily preceded by a sojourn in the never-in-labor force state.

Use of a delayed renewal-process specification has a very important consequence for estimation: we now condition on the state currently occupied. In Baydar's model, observations for individuals who have never moved would be discarded. The dummy variable d would always have the value 1. One might object that this approach does not use all the available sample information. However, it does use all the information that is relevant to the process being modelled, namely, the process of times **between** residential moves. Note that when a delayed renewal model is postulated, the process being modelled always begins with a renewal.

A third point relating to model specification concerns Baydar's correction for censoring. Her equation (7) is an expression for the joint probability of two events: backwards recurrence time, **given** survivorship, $= u$; and survivorship of censoring risks from birth to the survey date. The marginal probabilities of these two events are assumed to be independent. Note that the estimated values of the parameters of the likelihood function are unaffected by the presence of the term $S_c(x)$, as can readily be seen by examining the gradient of the log-likelihood implied by (7).

A less restrictive formulation would be one in which $S(u)$ is regarded as the joint survivor function for a number of competing risks; this implies, however, that $f(t) = -\frac{\partial}{\partial t} S(t)$ is the density of times from a particular event (eg. a residential move) to the next event **of any type**. The difference between the two approaches,

In the present substantive context, is captured by the phrases "duration of residence prior to changing addresses" and "duration of stay at a single address". The latter may be suitable in many applications. This approach is more general, since it admits the possibility that the competing risks are only **conditionally** independent (conditional, for example, on a single fixed heterogeneity term).

Baydar's summary repeats a statement found in Allison (1985), namely that the renewal-theoretic approach to modelling backwards recurrence time data requires that "...the event being modelled is a repetitive event...". I wonder, however, if it is not sufficient that the event merely coincide with entry into a **transient** state. For this more general condition to have any meaning, we must be talking about a multistate process, upon which more than one renewal process may be defined. In a J-state Markov renewal process, there exists a renewal function for successive occurrences of each of the J events. Consider the following stochastic process:

state space = {0 = parity 0; 1 = parity 1; 2 = parity 2; d = dead};

possible events = {own birth; birth of first child; birth of second child; own death};

kernel = {$Q(i,j,t)$},

where

$$\frac{d}{dt}Q(0,1,t) = b_0(t)\exp(-\int_0^t [b_0(x) + d_0(x)]dx) \ ,$$

$$\frac{d}{dt}Q(0,d,t) = d_0(t)\exp(-\int_0^t [b_0(x) + d_0(x)]dx) \ ,$$

$$\frac{d}{dt}Q(1,2,t) = b_1(t)\exp(-\int_0^t [b_1(x) + d_1(x)]dx) \ ,$$

$$\frac{d}{dt}Q(1,d,t) = d_1(t)\exp(-\int_0^t [b_1(x) + d_1(x)]dx) \ ,$$

and

$$\frac{d}{dt}Q(2,d,t) = d_2(t)\exp(-\int_0^t d_2(x)dx) \ ;$$

all other $Q(i,j,t) = 0$ for all t. In other words fertility risks (b_0 and b_1) and death risks (d_0, d_1, and d_2) are defined with respect to duration of current status occupancy only, and not age; this is not a very satisfying demographic model but

does serve to illustrate the point. And the point is that there is a renewal function for the event "first birth", even though the event cannot be repeated. The renewal density for the event "birth of first child", in this rather trivial model, is in fact

$$b_0(x)\exp(-\int_0^x [b_0(z) + \alpha_0(z)]dz) \quad.$$

Backwards recurrence time estimation could proceed with a sample recording current parity and duration of current parity status; if we selected only respondents currently at parity 1, we could estimate the distribution of times from first birth to second births or death, in a stable population.

I have one substantive point concerning Baydar's model specification. In particular, she specifies a mixing distribution for the marginal failure-time distribution [see (18) above], then introduces covariates in a standard proportional-effects way. The mixing distribution is justified on the familiar—and altogether reasonable—grounds that observed residential mobility is generated by a mixture of the behavior of "movers" and "stayers". In the mixture model the proportions of each are represented by the unknown parameters δ_1 and δ_2 ($=1 - \delta_1$). It seems rather odd to adopt a specification whereby the covariate effects operate on the weighted sum of the hazards of the two groups; that, in other words, the effect of a change in some X is the same regardless of whether one is a "mover" or a "stayer". A superior model, I would think, is one in which the effect of a covariate is to alter the chances that an individual is a "mover" or a "stayer". For example, we might suppose that the probability that someone is a mover is given by a logistic expression,

$$pr\,(i \text{ is a mover}) = pr\,(\delta_{1i} = 1) = 1 - \exp(-X_i B) \quad;$$

with the complementary probability for being a stayer. Let $h_1(t)$ and $h_2(t)$ represent the hazards of movers and stayers, respectively; then the survivor function for residential moves is

$$\exp-[\int_0^t h_1(z)dz - \exp(X\beta) \int_0^t h_1(z)dz + \exp(X\beta) \int_0^t h_2(z)dz] \quad;$$

this is a tractable expression, and is somewhat different from that appearing in Baydar's model.

To conclude, I would repeat that Baydar has made a useful contribution to a developing literature, one dealing with estimation and modelling issues that can become exceedingly complex. Baydar (like Allison) formulates a model in terms of the trajectory of a renewal process defined on an individual's lifetime; this leads, under assumptions that would be considered necessary in most applications to human behavior, to estimating equations that cannot be expressed in closed form; these in turn require difficult, messy, and possibly unstable and/or inaccurate computational techniques. I have suggested an alternative motivation, based upon the assumption that the population being sampled is stable, which leads to more tractable estimating equations; examples of applications in this spirit can be found in several labor economics papers (cited above), as well as in Greenberg and Wolf (1986). I have also suggested that it is not necessary to be dealing with repeatable events in order to analyze backwards recurrence-times data, appealing to the more general theory of Markov renewal processes.

As a final note, I think it worth mentioning that there is a middle ground between what has so far been treated as polar empirical situations: access to complete event-history data, and access strictly to backwards recurrence-time data. This middle ground, which may frequently be attainable with some digging, is having access to a combination of backwards recurrence-time data and the empirical renewal density. For example, the prototypical labor economist's problem--data on the duration of unemployment spells in progress, collected from those currently unemployed--could be solved more satisfactorily if, in addition, we had access to additional data giving counts of the newly-unemployed, by week, over time. Or, a duration-of-marriage model might be estimated by combining survey data on the duration of current marital status (backwards recurrence-time data) with counts, from another source (eg., vital records or population registers), of marriages by year. The strength of this approach is that it allows us to relax some stability assumptions; we must still assume stability, over time, of the survivor function for time since the last event in question, but need not assume that the process leading up to the last such event is stable over time. The problem with the approach is that it is frequently quite difficult to assemble a data series representing the empirical renewal density (or, more properly, the empirical renewal function) for the appropriate population. But, as Baydar himself points out, and as I imagine most of us would agree, there are bound to be problems in the analysis of backwards recurrence-time data.

REFERENCES

Allison, Paul D. (1985) Survival analysis of backwards recurrence times. *Journal of the American Statistical Association* 80:315-322.

Cinlar, Erhan (1975) *Introduction to Stochastic Processes*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Cox, David R. (1967) *Renewal Theory*. London: Methuen.

Flinn, Christopher J. (1986) Econometric analysis of CPS-type unemployment data. *Journal of Human Resources* (forthcoming).

Greenberg, David and Douglas A. Wolf (1986) The dynamics of welfare fraud: An econometric duration model in discrete time. *Journal of Human Resources* (forthcoming).

Kaitz, Hyman B. (1970) Analyzing the length of spells of unemployment. *Monthly Labor Review* 93 (November):11-20.

Nickell, Stephen (1979) Estimating the probability of leaving unemployment. *Econometrica* 47 (September):1249-66.

Salant, Stephen W. (1977) Search theory and duration data: A theory of sorts. *Quarterly Journal of Economics* 91 (February):39-57.

Discussion of Jan M. Hoem, Bo Rennermalm, and Randi Selmer: "Restriction biases in the analysis of births and marriages to cohabiting women from data on the most recent conjugal union only"

Rolf Ziegler

I will first give a brief comment and then raise three questions.

1. I want to stress again - as Jan Hoem did too - that there is no basis for generalizing the finding that the biases due to on observational plan like the Danish - i.e. data exist on the most recent conjugal union only - will not cover up or mask general trends among cohorts.

The false omissions of and partially incorrect information about cohabitational experiences lead, to an unknown degree, both to over- and underestimation of real decreases in rates of marriage or childbirth. Jan Hoem has stressed that fact in his paper and I just want to underline it.
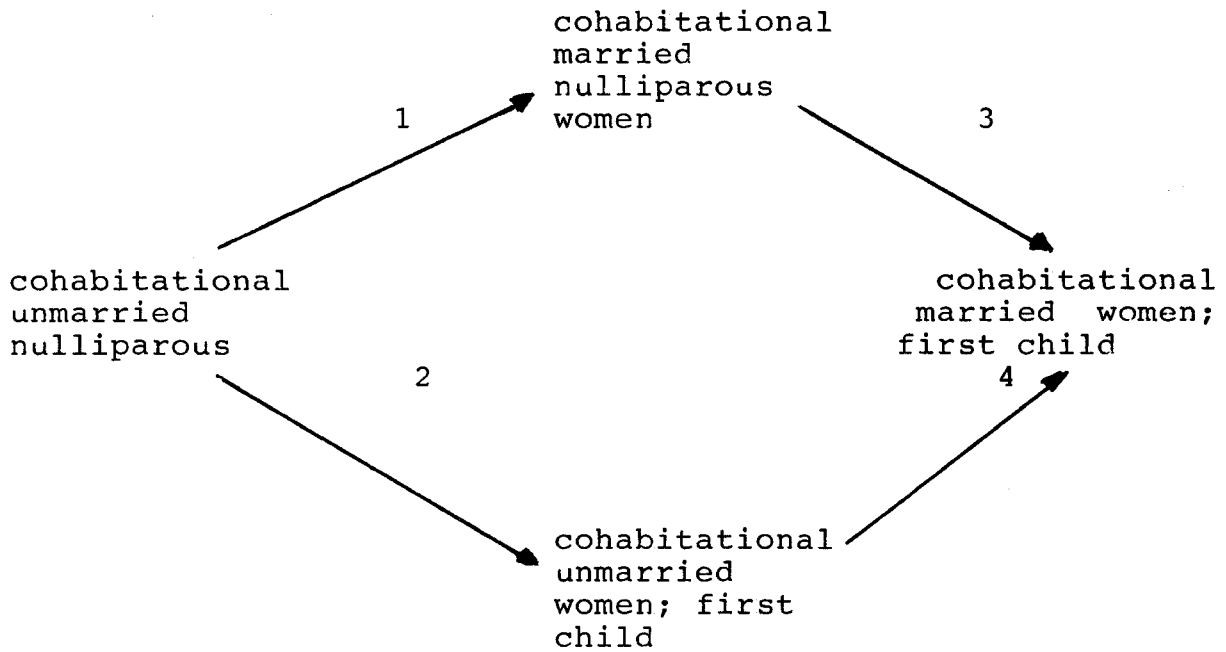
The omissions due to lack of information are about those women who have had cohabitational experience at age below 25 while unmarried and nulliparous but which
- either stopped before the time of interviewing
- or doesn't relate to the present cohabitational partner or (former) spouse.

The information gained from unmarried women about their present conjugal union or from married, separated or divorced women about any premarital cohabitational experience with their (former) spouse, may partially be incorrect if there has been an even earlier conjugal union.

My questions relate to the kind of analysis and not to possible errors of design.

2. As I understand, Jan Hoem analyzed the following competing risk model (transitions 1 and 2).

```
                        cohabitational
                        married
                        nulliparous
              1         women            3
```

```
cohabitational                          cohabitational
unmarried                               married  women;
nulliparous                             first child
              2                                  4
```

```
                        cohabitational
                        unmarried
                        women; first
                        child
```

It would be interesting to know how rates have changed if one includes the other two transition possibilities (3 and 4).

3. Wouldn't it be more adequate to consider the time when a woman becomes aware of her being pregnant as the decisive event in leading to marriage or not? This is of course what has been done in the paper by Angelika Toelke yesterday. I realize that one should then include abortion as another competing risk.

4. There are I think two different substantive questions both of them equally interesting. The first is: Is there a change in attitudes towards conjugal unions among unmarried couples? From all we know - including the present study - a very decisive change has taken place among younger cohorts.
The other question, however, is slightly different: Have the time-dependent risks becoming married changed for people who are going steady? If one takes into account the duration of engagements, waiting times until becoming married seem to have been rather long among earlier cohorts.
Of course there are obvious differences between an engage-

ment and cohabitation. But if one talks about casualness of relationships among men and women they are very similar in the following sense: engagement and cohabitation are public signals of the (temporary?) withdrawal from the marriage market.

Discussion of Reinhard Hujer and Hilmar Schneider: "Unemployment duration as a function of individual characteristics and economic trends"

Olaf Hübler

(1) The unemployment duration analysis is one of the tra-
ditional economic applications in event history analy-
sis. There exist several studies of this problem,
especially from the United States and Great Britain,
for example Lancaster (1979), Lancaster/Nickell (1980),
Heckman/Borjas (1980), Nickell (1979), Flinn/Heckman
(1982), Lynch (1985), Moffitt (1985), Addison/Portugal
(1986a). In Germany the problems of unemployment dura-
tion were only considered with aggregated data until
now with the exception of Büchtemann/Brasche (1985),
who based their analysis on individual data, but only
used cross tables and frequency distributions to de-
scribe the problems.

Hujer/Schneider have presented a very interesting
paper on unemployment duration in Germany by using
the instruments of event history analysis. Especially
four issues of their study are important:

- They give some theoretical explanations of unemployment duration,
- they investigate whether "Arbeitslosengeld" (unemployment benefits) has an effect on the unemployment duration
- they introduce global, aggregated indicators to explain individual unemployment duration
- they distinguish between transition probabilities from unemployment to employment and from unemployment to "out of work".

In the following I shall supplement some of these points and compare Hujer/Schneiders' results with those of other investigations.

(2) Most empirical studies of unemployment duration are based theoretically on job search theory. The key relationship is that between the reservation wage and the probability of leaving unemployment. Utilizing a standard search model we know that the probability that an unemployed worker will become re-employed is the product of three probabilities- the probability of searching, the probability of receiving a job offer, and the probability of accepting the job offer. The first probability is determined by the reservation wage and the unemployment benefits, the second depends on personal characteristics and global demand conditions, and the third results from the reservation wage and the distribution of wage offers. From this view it is obvious to include global demand indicators, personal indicators which are relevant to productivity, and the "Arbeitslosengeld" to explain the probability of leaving the unemployment state.

The reservation wage cannot be observed. There exist different methods of approximation. In the simplest

search models only the search costs, the wage-offer
distribution, and possibly a discount factor determine
the reservation wage and hence the unemployment dura-
tion. In more general models the reservation wage is
determined by the arrival time of offers, attitudes of
risk, moving costs, past experiences, human capital,
age, and other personal characteristics. Search costs
are measured in terms of travel costs and unemployment
benefits. Most of these variables can be found in the
study of Hujer/Schneider. Panel data sometimes contain
a question attempting direct measurement of the reser-
vation wage. The question "What is the lowest amount
per week you are prepared to accept in a new job", for
example, comes as close as any question to an enquiry
about a reservation wage. But Stephanson (1976) has
noted that there is an important difference between
the reservation wage concept of search theory and the
"asking wage" notion derived from sample survey.

(3) Usual personal characteristics which are introduced in
    unemployment duration analysis are age, sex, schooling,
    tenure, experiences, race. Recently Addison/Portugal
    (1986a) following Folbre/Leighton/Roderick (1984) have
    emphasized and empirically shown in an event history
    model that advance notification of plant closing sig-
    nificantly lowers the unemployment duration resulting
    from such dislocation. Normally no data exist to con-
    sider this effect. But it can be important in the sense
    of unobserved heterogeneity. I mention this variable,
    too, because this is no personal characteristic and no
    global indicator. I suggest to supplement the mentioned
    unemployment duration determinants by characteristics
    which describe the last firm of the unemployed per-
    sons. I suppose that the firm size and the sector are
    relevant, too.

Benefits from unemployment insurance are very often
introduced as a determinant of unemployment duration.
From search theory the expected sign of this relation-
ship is obvious. The benefits have a positive effect
on unemployment duration. Benefits reduce the search
costs and therefore increase the reservation wage. But
empirical results are ambiguous. If the effect of unem-
ployment benefits or the replacement ratio with unem-
ployment benefits as numerator and the wage before un-
employment as denumerator are used results normally
agree with search theory,see for example Feinberg
(1977),Chapin (1971), Ehrenberg/Oaxaca (1976),Lancaster
(1979), Lynch (1985), Addison/Portugal(1986a). But if the
effect of maximum duration of paying benefits from unemploy-
ment insurance are examined the results are mixed.
Studies by Clark/Summers (1983), Ehrenberg/Oaxaca
(1976),Kiefer/Neumann (1979) yield insignificant coef-
ficients. Moffitt (1985) shows that the result depends
on the specification. There exist only few studies
which agree with Hujer/Schneiders' result that the
transition rate of leaving the unemployment state is
larger for persons with entitlement to benefits from
unemployment insurance than for those without entitle-
ment. Some years ago there was a controversy between
König (1978) and Egle/Karr (1980) on this issue. I
think it is very difficult to decide this question
unambiguously.

As an explanation of their result Hujer/Schneider
emphasize that the majority of the population without
entitlement consists of entry-level persons or persons
with a longer interruption of labour market participa-
tion. This is an aspect of lagged duration dependence.
The argument of Hujer/Schneider can be supplemented.
Persons with high "Arbeitslosengeld" or with entitle -
ment to "Arbeitslosengeld"have earned more money than

others in former times, which is a screening device for
firms that these persons are more productive than others,
that they have more experience. It is also possible
that persons with higher "Arbeitslosengeld" are more
intelligent, have a more efficient search strategy,
and thus have more opportunities to find a new job.
Perhaps they are more motivated to find a new job be-
cause their absolute difference between former wage
or the expected net earnings and "Arbeitslosengeld"
is larger than that of other persons. Some of these
hypotheses can be examined if there exist data to
introduce the suitable variables,but other effects
remain unobserved. I am not sure whether Hujer/Schnei-
ders' result is a specific one for Germany or whether
the crude measurement of unemployment compensation is
relevant for their result and the results can be the
same in other countries if the same specification is
used. Further studies are necessary.

(4) My next comment pertains to the global indicators.
Most empirical studies of unemployment duration with
individual data do not use aggregated variables as
regressors or they include only the local or national
rate of unemployment (Burdett et.al.1985, Lynch 1985,
Addison/Portugal 1986a). Results of latter studies
normally indicate that the transition rate from un-
employment to employment falls significantly if the
aggregated unemployment rate increases (Lancaster 1979,
Moffitt 1935) and correspond to those of Hujer/Schnei-
der. But in Heckman/Borjas (1980) and Flinn/Heckman
(1982) the coefficient of national unemployment rate
is not significant.

From a theoretical view it is sensible to include more
global indicators than the unemployment rate into the

model, but they are often highly correlated and sometimes there exist opposite effects so that the signs of the coefficients are ambiguous. Take the number of vacancies or the vacancy rate (Barron 1975). This variable affects the duration of unemployment in at least three ways. In the first, and the most direct, a fall in the vacancy rate makes a job contact less likely and ceteris paribus prolongs the expected duration of unemployment. In the second, this direct effect is couteracted by a downward revision of the reservation wage in the light of lower returns to search, thus lowering the expected duration. In the third, a lower vacancy rate may, in general cyclical downswing, be accompanied by lower wage offers, so that the wage distribution shifts to the left. If this shift is not perceived searchers will find fewer jobs acceptable and the expected duration of unemployment increases. Axelsson/Löfgren (1977) find the first effect dominant so that a lower vacancy rate tends to increase the duration. But it should be mentioned as Kühl (197o) has stressed that the number of vacancies is no valid indicator of labour demand.

Another problem with labour demand indicators as the number of vacancies is that they do not affect all groups of unemployed in the same way. For example, if we had a crisis and the demand begins to increase it is possible that the duration of unemployed, highly educated persons with experience shortens while the unemployment duration of unexperienced still increases. I believe interaction variables between personal characteristics and global indicators can partially catch these counteracting effects. In other words, the coefficients of labour demand (β_D) vary systematically with the individual experience (EX):

$$\beta_D = a_o + a_1 EX + a_2 EXSQ \qquad (1)$$

where a_i are fixed coefficients (i=o,1,2) and EXSQ is
square of experience. Another example is the discouraged
and the added worker effect if labour demand falls. In
families in which the women are normally out of work
the supply of labour of these women and usually the
duration of unemployment increase if the husband is un-
employed. In other families in which the women are un-
employed for a long time they are possibly discouraged
after some time and leave the labour force. Using data
of official statistics the unemployment duration of
this group does not increase with the reduction of
labour demand.

(5) As we have seen by the discouraged and added worker
effect it is sensible - as Hujer/Schneider have done
it - to separate between transition probabilities from
unemployment to employment and from unemployment to
the non-active population. It is obvious from the re-
sults that there are differences. Especially, labour
market indicators affect the first probability but not
the latter. But in other cases - for example whether
the transition probability from unemployment to sector
A is the same as to sector B - it may be useful to
test whether the differences are significant following
the method of Flinn/Heckman (1983). They compare the
three state generalization of a two state model. If we
use the following symbols

e-employment, u-unemployment, o-out of the labour force
$P_j(t)$-probability that state j is occupied at time t
$\dot{P}_j(t)$-instantaneous rate of change of $P_j(t)$
h_{ij}-hazard rate for exit to state j from state i,
we can compare the three state Markov model by the

Kolmogorov-Chapman conditions

$$
\begin{bmatrix} \dot{P}_u(t) \\ \dot{P}_o(t) \\ \dot{P}_e(t) \end{bmatrix} = \begin{bmatrix} -(h_{ue}+h_{uo}) & h_{ou} & h_{eu} \\ h_{uo} & -(h_{ou}+h_{oe}) & h_{eo} \\ h_{ue} & h_{oe} & -(h_{eu}+h_{eo}) \end{bmatrix} \begin{bmatrix} P_u(t) \\ P_o(t) \\ P_e(t) \end{bmatrix} \quad (2)
$$

with the aggregated two state Markov model

$$
\begin{bmatrix} \dot{P}_u(t) \\ \dot{P}_n(t) \end{bmatrix} = \begin{bmatrix} -h_{un} & h_{nu} \\ h_{un} & -h_{nu} \end{bmatrix} \begin{bmatrix} P_u(t) \\ P_n(t) \end{bmatrix} \quad (3)
$$

We have to see whether state o and e can be aggregated to state n. If this condition holds, then follows

$$
P_n(t) = P_o(t) + P_e(t) \quad . \quad (4)
$$

A necessary condition that the three state system is equivalent to the two state system is that the two hazard rate matrices in (2) and (3) have the same rank and this has to be 1 because the rank of the two state system is 1. A necessary and sufficient condition can be derived if we substitute the condition of postulated equivalence (4) into the three state model and compare it with the two state model. Then we have

$$
-h_{nu}P_n(t) = -(h_{ou}+h_{oe})P_o(t)+h_{eo}P_e(t)+h_{oe}P_o(t)-(h_{eu}+h_{eo})P_e(t)
$$
$$
= -h_{ou}P_o(t)-h_{eu}P_e(t) \quad . \quad (5)
$$

It is easily seen that condition (5) is fulfilled if

$$
h_{nu} = h_{ou} = h_{eu} \quad (6)
$$

because

$$
-h_{nu}(P_o(t)+P_e(t)) = -h_{ou}P_o(t)-h_{eu}P_e(t)+(h_{ou}-h_{nu})P_o(t)+(h_{eu}-h_{nu})P_e(t) \quad .
$$
$$
(7)
$$

In order to test the hypothesis $h_{nu}=h_{ou}=h_{eu}$ the parameter estimates of the three state unrestricted and restricted model have to be compared by the likelihood ratio test. The restriction is that the coefficients β_{ou} from

$$h_{ou} = \exp(x'_{ou}\beta_{ou}) \qquad\qquad (8)$$

are the same as β_{eu} from

$$h_{eu} = \exp(x'_{eu}\beta_{eu}) \ . \qquad\qquad (9)$$

If $-2\left[\ln L_r - \ln L_u\right] = -2\left[\ln L(\beta_r) - \ln L(\beta_{ou}) - \ln L(\beta_{eu})\right] > \chi^2_{l,1-\alpha}$ the hypothesis of a two state system is rejected where L_r is the likelihood of the restricted model and L_u the likelihood of the unrestricted model, l is the difference between the degrees of freedom of the two models, the number of imposed restrictions.

(6) An important point in unemployment duration models is the question of state dependence. Heckman/Borjas (1980) and Heckman (1981) have distinguished between three kinds of state dependence:
- duration dependence
- lagged duration dependence
- occurence dependence.

The first indicates the duration in the present state in t, the second the duration in former states and the third the number of former states. In other words, working history is relevant for the expected rest duration of a state.

Different methods exist to incorparate the duration dependence into event history models. Hujer/Schneider have discussed some of these approaches. One specification has received much attention in the literature, the proportional hazard model (see Cox 1972). This

means

$$h(t) = \psi_1(t)\psi_2(x) .$$ (1o)

The natural form for ψ_2 is the exponential

$$\psi_2(x) = \exp(x'\beta) .$$ (11)

For $\psi_1(t)$ it is usually assumed that

$$\psi_1(t) = \alpha t^{\alpha-1} \qquad \alpha > o$$ (12)

so that we have a Weibull model. If $\alpha = 1$ the hazard rate is constant. From

$$d\psi_1/dt = (\alpha-1)\alpha t^{\alpha-2} = (\alpha-1)\psi_1/t$$ (13)

where $(\alpha-1)$ can be interpreted as time elasticity of the hazard rate, it follows that $d\psi_1/dt \gtrless o$ if $\alpha \gtrless 1$. In other words, if the hazard rate is increasing the subsequent expected duration of a spell of unemployment will be smaller the longer the individual is unemployed. Then we have positive duration dependence. Job search theory predicts this kind of duration dependence because the reservation wage falls if the spell of unemployment lengthens. But, on the other hand, if employers consider unemployment as a signal of low potential productivity then we have negative duration dependence. Most empirical studies indicate negative duration dependence (Lynch 1985, Lancaster 1979, Flinn/ Heckman 1982). This means, the longer the current spell of unemployment the more difficult it is to become re- employed. But it is possible that this effect is spurious if there exists unobserved heterogeneity, because un- controlled variables bias estimated hazards toward negative duration dependence (Heckman/Singer 1985, p. 53). This is easily seen because persons with unob- served characteristics which are a signal of high pro-

ductivity leave the unemployment state first and there-
fore only persons with bad signals stay unemployed and
have a poor chance for re-employment.

Another problem exists if we have time-varying variables
(x(t)) so that our model is

$$\ln h(t) = x(t)'\beta + ((t^\lambda - 1)/\lambda)\gamma + \theta(t) \tag{14}$$

with a Box-Cox conditional hazard model introduced by
Flinn/Heckman (1982) where $\theta(t)$ is the term of unob-
served heterogeneity. The formulation of the hazard
rate h(t) in (14) contains special cases as the Weibull
model ($\lambda=0$). If $x(t)=(t^\lambda-1)/\lambda$ it is impossible to separate
the effect of time-varying variables from duration de-
pendence. Then we have the classical problem of exact
multicollinearity. The dilemma is that failure to control
for time-varying regressor variables may mislead, but
introducing such variables may create an identification
problem.

Misspecification of the error distribution can induce
great sensitivity of the estimates. Heckman/Singer
(1982) have shown strong differences in the estimates
and the inference when they have used normal, lognormal,
or gamma heterogeneity. A study of Addison/Portugal
(1986b) only partially supports these results. They
compare lognormal, gamma, Weibull, reciproce Weibull,
exponential and log logistic distribution. It can be
seen that neither the sign nor the significance of the
parameters are sensitive to distributional assumptions.
However, when we turn to consider the magnitude of the
parameters dramatic changes are observed for some
variables.

Lagged duration and occurence dependence is rarely
studied. Büchtemann/Brasche (1985) find that in Germany

persons with "short" initial spells (< 3 months) have
shorter duration of last unemployment spell (1979-82)
than persons with "long" initial spells of unemployment
(6 months and longer). But they only use cross tables.
Lynch (1985) has included the length of previous com-
pleted spells in a hazard model. Her results do not
show evidence of lagged duration dependence.

(7) Let me summarize my comment:

   (1) The reservation wage concept of search theory to
      explain unemployment duration can only partially
      satisfy.

   (2) It is sensible to supplement determinants of unem-
      ployment duration by characteristics of the last
      firm of unemployed persons.

   (3) Although the result for the "Arbeitslosengeld" by
      Hujer/Schneider is contrary to the results of other
      studies it can be explained well.

   (4) I suggest to include interactions between personal
      characteristics and global indicators into the un-
      employment duration model.

   (5) In Markov models it is possible to test whether
      two states as employment and out of labour force
      can be aggregated.

   (6) State dependence, especially duration dependence,
      should explicitly be considered although some
      methical problems follow.

References:

Addison, J.T. and P. Portugal (1986a), The Effect of Advance Notification of Plant Closings on Unemployment, mimeo., University of South Carolina.

Addison, J.T. and P. Portugal (1986b), On the Distributional Shape of Unemployment Duration, mimeo., University of South Carolina.

Axelsson, R. and K.G. Löfgren (1977), The Demand for Labour and Search Activity in the Swedish Labour Market, European Economic Review 9, 345-36o.

Barron, J.M. (1975), Search in the Labor Market and the Duration of Unemployment: Some Empirical Evidence, American Economic Review 65, 934-942.

Büchtemann, C.F. and U. Brasche (1985), Recurrent Unemployment. Longitudinal Evidence for the Federal Republic of Germany, SAMF-Arbeitspapier 1985-3.

Burdett, R., Kiefer, N.M. and S. Sunil (1985), Layoffs and Duration Dependence in a Model of Turnover, Journal of Econometrics 28, Annals 1985-1, 51-7o.

Chapin, G. (1971), Unemployment Insurance, Job Search and the Demand for Leisure, Western Economic 9, 1o2-1o7.

Clark, R. and L. Summers (1983), Unemployment Insurance and Labour Market Transitions, in: M.N. Baily (ed.), Workers, Jobs, and Inflation, Washington.

Cox, D. (1972), Regression Models and Life Tables, Journal of the Royal Statistical Society B 35, 187-22o.

Egle,F. and W. Karr (198o), Freiwillige Arbeitslosigkeit? Anmerkung zum Aufsatz von Heinz König, Kyklos 33, 142-152.

Ehrenberg, R.G. and R. Oaxaca (1976), Unemployment Insurance Duration of Unemployment and Subsequent Wage Gain, American Economic Review 66, 754-766.

Feinberg, R. M. (1977), Risk-aversion, Risk and the Duration of Unemployment, Review of Economics and Statistics 59, 264-271.

Flinn, C. and J. Heckman (1982), Models for the Analysis of Labor Force Dynamics, in: R. Bassmann and G. Rhodes, Advances in Econometrics 1, Greenwich, 35-95.

Flinn, C. and J. Heckman (1983),Are Unemployment and Out of the Labor Force Behaviorally Distinct Labor Force States? Journal of Labor Economics 1, 28-42.

Folbre ,N.R., Leighton, J.L. and M.R. Roderick (1984), Plant Closings and Their Regulation in Chaine, 1471-1482, Industrial and Labor Relations Review 37, 185-196.

Heckman, J. (1981), Heterogeneity and State Dependence, in: S. Rosen (ed.), Studies in Labor Markets, Chicago, 91-139.

Heckman, J. and G.J. Borjas (198o), Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence, Economica, 247-283.

Heckman, J. and B. Singer (1982), The Identification Problem in Econometric Models for Duration Data, in: W. Hildenbrand (ed.), Advances in Econometrics, Cambridge.

Heckman, J. and B. Singer (1985), Longitudinal Analysis of Labor Market Data, Cambridge.

Kiefer, N.M. and G.R. Neumann (1979), An Empirical Job Search Model with a Test of the Constant Reservation Wage Hypothesis, Journal of Political Economy 87, 89-1o7.

König, H. (1978), Zur Dauer der Arbeitslosigkeit: Ein Markov-Modell, Kyklos 31, 36-52.

Kühl, J. (197o), Zum Aussagewert der Statistik der offenen Stellen, Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 4, 419-449.

Lancaster, T. (1979), Econometric Methods for the Duration of Unemployment, Econometrica 47, 939-956.

Lancaster, T. and S. Nickell (198o), The Analysis of Re-employment Probabilities for the Unemployed,Journal of the Royal Statistical Society A, 141-165.

Lynch, L.M. (1985), State Dependence in Youth Unemployment: A Lost Generation? Journal of Econometrics 28, Annals 1985-1, 71-84.

Moffitt,R. (1985), Unemployment Insurance and the Distribution of Unemployment Spell, Journal of Econometrics 28, Annals 1985-1, 85-1o2.

Nickell, S. (1979), Estimating the Probability of Leaving Unemployment, Econometrica 47, 1249-1266.

Stephanson, S.P. (1976), The Economics of Youth Job Search Behavior, Review of Economics and Statistics 58, 1o4-111.

Kommentar zu H. J. Andreß: „Spezifikationsfehler in Regressionsmodellen für Übergangsraten"

Lothar Krempel

Andress's paper adresses two issues: (1) what the consequences are of ignoring systematic components when estimating rates and (2) whether the use of time-dependent models improves estimation under misspecification.

After a short introduction into rate models he presents in part two of his paper analytical terms for the effects which result when ignoring a systematic source of covariation (a second, in the model itself unspecified, variable). The reported bias becomes smaller with time. Its amount varies with the correlation of the unspecified and explicit component of the model.

Part three gives results for specification errors when using time dependent and non-parametric estimation procedures controlling the intercorrelation of explicit and unspecified components, the degree of censoring and the sample size.

The analytical results of part two and the corresponding simulations demonstrate how complicated consequences of misspecifications are - even when there is no correlation with the unspecified component.

Part three of the paper is motivated by the vague hope that the consequences of theoretical misspecification may be salvaged by more complicated models (time dependent or nonparametric functions) of the rate. The results show that there is no different efficiency of these models in reducing the effects of misspecification under any of the simulated conditions whereas correlation of explicit and unspecified components, censoring and sample size result in drastic effects.

The most serious and unpleasant result of this paper is that estimations are biased even when systematic and unspecified components are uncorrelated.

I will focus in my comment on the validity of these results: how reasonable are the assumptions under which they were derived and how important are they for the application of rate models ?

The paper itself presents four sources of estimation bias: misspecification, correlation, censoring and sample size. One way to discuss the reported results is

to question the variation of all four bias sources in
terms of their validity.

If we take censoring as an example to start with, we
find effects of censoring (part three) to be almost
indistinguishable in size and sign from effects reprted
for different sample size. One could argue that both
are almost the same because their effects always point
in the same direction. In my application practice I
found censoring to cause much more serious problems:

The underlying assumption of this simulation is that
all censored observations follow the same model as the
uncensored ones. While this is the the common
assumption to treat censored events in maximum-likeli-
hood estimators it may be too optimistic for some
social science applications: panel mortality, drop outs
etc. often lead to the suspicion that censored
observations are governed by a different process. As
long as we have theoretical arguments that censored and
uncensored observations are indicators of the same
process the results appear to be as they are presented
in this paper - if not it becomes worse. Thus the
results for censoring are obtained under relatively
favorable conditions.

Effects of missspecification are presented by
contrasting a strong positive with a strong negative
correlation. I argue that this not a realistic starting
point to evaluate effects of misspecification. It is
reasonable to expect, that theoretical knowledge will
guide us to avoid such extreme misspecifications.

The key argument which leads me to believe that conse-
quences for applications are less important than those
reported in this study is the common regression
assumption about the distribution of the error terms.
In regression analysis we assume that the expectation
of the error distribution is zero: there are serveral
unrelated errors (of different sign), so that their
individual contributions cancel out.

The paper shows us that we can misspecify models in
such a way that serious bias results when we employ the
currently available estimation procedures. Bias which
results in such a case is not 'captured' by common
estimators of time dependent models: there is no
'cheap' formal solution to the problem. From the two
ways to cope with unobserved heterogeneity I clearly
favor the substantial one.

If one favors the formal option, one has to leave the
problem of unobserved hetrogeneity to statisticians
until they create a 'superterm', which 'eats up what

we have left on our plate'. If a formal solution
exists, it will be most likely complicated and need
very strong assumptions to be applied, if possible at
all.

The substantial option is to elobarate on our models.
Theory should provide us with a catalog of covariates
which have to be controlled in any serious application.
My impression is, that we are still at the beginning of
theory-guided, exhaustive research with process models.
Any improvement will decrease specification errors.
Statistics for misspecification like the index
presented by Gerhard Arminger on this conference may
help to determine the point when our substantial
arguments have exhausted systematic components in our
data.

Veröffentlichungen der Reihe MATERIALIEN AUS DER BILDUNGSFORSCHUNG

Beim Max-Planck-Institut für Bildungsforschung erhältliche Bände

1 Christof Conrad
Schulsysteme im quantitativen Vergleich — Hamburg und
Westberlin.
Statistische Indikatoren für Demokratisierung und Moder-
nisierung im Schulwesen.
Berlin: Max-Planck-Institut für
Bildungsforschung 1972.   DM 10,—

2 Christiane Bierbaum
Die schwedische Schuldemokratie.
Ein Modell für die Schulreform in der Bundesrepublik? Ein
Bericht über Hintergründe, Verlauf und Ergebnisse der
Demokratisierung der Schule in Schweden.
Berlin: Max-Planck-Institut für
Bildungsforschung 1975.   DM 10,—

3 Frank Braun, Detlef Glowka, Klaus-Dieter Mende,
Peter Müller, Helga Thomas und Jürgen Zimmer
Schulreform und Gesellschaft.
Vergleichende Studie über die gesellschaftlichen Bedingungen
von Schulreformen in sieben europäischen Ländern.
Teil I und Teil II.
Berlin: Max-Planck-Institut für
Bildungsforschung 1975.

4 Heinrich Meulemann
Wortbedeutungsverständnis und Wortbedeutungsexplikation.
Eine empirische Analyse zweier Aspekte des Sprachverhaltens
und ihrer sozialen Determinanten im Rahmen der Theorie
der linguistischen Codes.
Teil I und Teil II.
Berlin: Max-Planck-Institut für
Bildungsforschung 1976.   DM 33,—

5 Helga Gripp
Zur Struktur ehelicher Interaktion.
Determinanten der Genese pathologischer Kommunikations-
formen in der Ehe und ihre Behandlung in der Therapie.
Eine Fallanalyse.
Zweite Auflage.
Berlin: Max-Planck-Institut für
Bildungsforschung 1978.   DM 22,—

6 Helmut Köhler
Daten zur Situation der Hauptschule in Berlin (West).
Berlin: Max-Planck-Institut für
Bildungsforschung 1976.   DM 6,—

7 Yvonne Schütze
Innerfamiliale Kommunikation und kindliche Psyche.
Eine exemplarische Analyse der Kommunikations- und
Rollenstruktur zweier Familien.
Zweite, korrigierte Auflage.
Berlin: Max-Planck-Institut für
Bildungsforschung 1978.   DM 27,—

8 Helmut Köhler
Quellen der Bildungsstatistik.
Eine kommentierte Zusammenstellung statistischer
Veröffentlichungen.
Berlin: Max-Planck-Institut für
Bildungsforschung 1977.   DM 8,—

9 Ulrich W. Bamberg
Leistungsbezogene Persönlichkeitsmerkmale gelernter
Maschinenschlosser.
Berlin: Max-Planck-Institut für
Bildungsforschung 1977.   DM 10,—

10 Peter Siewert und Helmut Köhler
Grundschulfinanzierung und Grundschulpolitik.
Aufgaben und Lastenverteilung im Primärbereich.
Berlin: Max-Planck-Institut für
Bildungsforschung 1977.   DM 8,—

11 Barbara Hegelheimer
Berufsqualifikation und Berufschancen von Frauen in der
Bundesrepublik Deutschland.
Berlin: Max-Planck-Institut für
Bildungsforschung 1977.   DM 14,—

12 Wolfgang Lempert
Untersuchungen zum Sozialisationspotential gesellschaft-
licher Arbeit.
Ein Bericht.
Berlin: Max-Planck-Institut für
Bildungsforschung 1977.   DM 11,—

13 Helmut Köhler
Der relative Schul- und Hochschulbesuch in der Bundes-
republik Deutschland 1952 bis 1975.
Ein Indikator für die Entwicklung des Bildungswesens.
Berlin: Max-Planck-Institut für
Bildungsforschung 1978.   DM 10,—

14 Wolfgang Lempert, Ernst Hoff und Lothar Lappe
Konzeptionen zur Analyse der Sozialisation durch Arbeit.
Theoretische Vorstudien für eine empirische Untersuchung.
Zweite Auflage.
Berlin: Max-Planck-Institut für
Bildungsforschung 1980.   DM 29,—

15 Marianne Müller-Brettel
Die Diskussion der Arbeitslehre 1964–1979.
Eine annotierte Bibliographie.
Berlin: Max-Planck-Institut für
Bildungsforschung 1979.   DM 8,—

16 Klaus Stanjek
Die Entwicklung des menschlichen Besitzverhaltens.
Berlin: Max-Planck-Institut für
Bildungsforschung 1980.   DM 13,—

17 Claudia von Grote
Die Bedeutung der soziolinguistischen Kodes für die kommu-
nikativen Fähigkeiten eines Sprechers.
Eine empirische Analyse der objektiven Kommunikations-
effizienz schichtenspezifischer Sprechweisen in variierenden
situativen Kontexten.
Berlin: Max-Planck-Institut für
Bildungsforschung 1980.   DM 28,—

18 Ulf Homann
Die Diskussion der Schulpflichtverlängerung in der Bundes-
republik Deutschland 1949 bis 1979.
Eine annotierte Bibliographie.
Berlin: Max-Planck-Institut für
Bildungsforschung 1981.   DM 10,—

19 Helmut Köhler und Luitgard Trommer
Quellen der Bildungsstatistik auf Länder- und Gemeinde-
ebene.
Berlin: Max-Planck-Institut für
Bildungsforschung 1981.   DM 12,—

20 Wilke Thomssen
Verarbeitung von beruflichen und betrieblichen Erfahrungen.
Gruppendiskussionen mit Befragten der Maschinenschlosser-
studie über ausgewählte Ergebnisse dieser Studie.
Berlin: Max-Planck-Institut für
Bildungsforschung 1981.   DM 20,—

21 Karl Anders
Von Worten zur Syntax: Spracherwerb im Dialog.
Berlin: Max-Planck-Institut für
Bildungsforschung 1982.   DM 19,—

22 Siegfried Reuss
Die Verwirklichung der Vernunft.
Hegels emanzipatorisch-affirmative Bildungstheorie.
Berlin: Max-Planck-Institut für
Bildungsforschung 1982.   DM 52,—

23 Jürgen Peter Hess
Empirische Sozialforschung und automatisierte Daten-
verarbeitung.
Berlin: Max-Planck-Institut für
Bildungsforschung 1982.   DM 17,–

24 Ernst Hoff, Lothar Lappe und Wolfgang Lempert
Methoden zur Untersuchung der Sozialisation junger Fach-
arbeiter.
Teil I und Teil II.
Berlin: Max-Planck-Institut für
Bildungsforschung 1983.   DM 48,–

25 Wolfgang Hoebig
Bedürftigkeit – Entfremdung der Bedürfnisse im
Kapitalismus.
Berlin: Max-Planck-Institut für
Bildungsforschung 1984.   DM 37,–

26 Gundel Schümer
Daten zur Entwicklung der Sekundarstufe I in Berlin (West).
Berlin: Max-Planck-Institut für
Bildungsforschung 1985.   DM 19,–

27 Hans-Uwe Hohner
Kontrollbewußtsein und berufliche Restriktivität.
Entwicklung und empirische Erprobung eines integrativen
Modells.
Berlin: Max-Planck-Institut für
Bildungsforschung 1985.   DM 49,–

28 Ingeborg Tölke
Ein dynamisches Schätzverfahren für latente Variablen in
Zeitreihenanalysen.
Berlin: Max-Planck-Institut für
Bildungsforschung 1986.   DM 17,–

29 Ulrich Trommer
Aufwendungen für Forschung und Entwicklung in der
Bundesrepublik Deutschland 1965–1983.
Theoretische und empirisch-statistische Probleme.
Berlin: Max-Planck-Institut für
Bildungsforschung 1987.   DM 32,–

30 Karl Ulrich Mayer and Nancy Brandon Tuma (Eds.)
Applications of Event History Analysis in Life Course
Research.
Berlin: Max-Planck-Institut für
Bildungsforschung 1987.   DM 48,–

Nicht über den Buchhandel erhältliche

Schriftenreihen aus dem

Max-Planck-Institut für Bildungsforschung

– Beiträge aus dem Forschungsbereich Entwicklung und
Sozialisation bzw. Schule und Unterricht.

– Literatur-Informationen aus der Bildungsforschung.

Veröffentlichungen der Reihe STUDIEN UND BERICHTE des Max-Planck-Instituts für Bildungsforschung

Im Buchhandel erhältliche Bände (über den Verlag Klett-Cotta zu beziehen)

38 Sigurjón Björnsson and Wolfgang Edelstein
in collaboration with Kurt Kreppner
Explorations in Social Inequality.
Stratification Dynamics in Social and Individual
Development in Iceland.
172 S. Erschienen 1977.
ISBN 3-12-98242 0-0

39 Reinhard Franzke
Berufsausbildung und Arbeitsmarkt.
Funktionen und Probleme des „dualen Systems".
211 S. Erschienen 1978.
ISBN 3-12-98243 0-8

40 Beate Krais
Qualifikation und technischer Fortschritt.
Eine Untersuchung über Entwicklungen in der industriellen
Produktion.
143 S. Erschienen 1979.
ISBN 3-12-98244 0-5

41 Jürgen Baumert
in Zusammenarbeit mit Diether Hopf
Curriculumentwicklung und Lehrerfortbildung für die
Berliner Gesamtschulen.
Ergebnisse von Lehrerbefragungen zur curriculumbezogenen
Fortbildung und zur Rekrutierung von Gesamtschullehrern.
167 S. Erschienen 1980.
ISBN 3-12-98245 0-2

42 Manfred Auwärter
Sprachgebrauch in Abhängigkeit von Merkmalen der Sprecher
und der Sprechsituation.
Eine soziolinguistische Untersuchung.
365 S. Erschienen 1982.
ISBN 3-12-98246 0-X

43 Eberhard Schröder
Entwicklungssequenzen konkreter Operationen: Eine
empirische Untersuchung individueller Entwicklungsverläufe
der Kognition.
112 S. Erschienen 1986.
ISBN 3-608-98247-7

Vergriffene Bände (Restexemplare teilweise noch im Max-Planck-Institut für Bildungsforschung erhältlich)

1 Marianne von Rundstedt
Die Studienförderung in Frankreich in den Jahren
1950–1962.
108 S. Erschienen 1965.

2 Fritz Joachim Weiß
Entwicklungen im Besuch berufsbildender Schulen in den
Ländern der Bundesrepublik Deutschland 1957 bis 1963.
108 S. Erschienen 1965.

3 Lothar Krappmann
Die Zusammensetzung des Lehrkörpers an den Pädagogischen
Hochschulen und entsprechenden Einrichtungen.
Wintersemester 1964/65.
155 S. Erschienen 1966.

4 Klaus Herzog
Das Arbeiten mit Kostenlimits im englischen Schulbau.
Ministry of Education, London.
Kostenstudie.
191 S. Erschienen 1965.

5 Marianne von Rundstedt
Die Studienförderung in Belgien 1950 bis 1963.
115 S. Erschienen 1966.

6 Gerhard Kath, Christoph Oehler und Roland Reichwein
Studienweg und Studienerfolg.
Eine Untersuchung über Verlauf und Dauer des Studiums von
2.000 Studienanfängern des Sommersemesters 1957 in
Berlin, Bonn, Frankfurt/Main und Mannheim.
Mit einem Vorwort von Dietrich Goldschmidt.
228 S. Erschienen 1966.

7 Wolfgang Lempert
Die Konzentration der Lehrlinge auf Lehrberufe in der
Bundesrepublik Deutschland, in der Schweiz und in
Frankreich 1950 bis 1963.
Eine statistische Untersuchung.
98 S. Erschienen 1966.

8 Rosemarie Nave-Herz
Vorberuflicher Unterricht in Europa und Nordamerika — eine
Übersicht.
Eingeleitet von Wolfgang Lempert.
152 S. Erschienen 1966.

9A Klaus Hüfner
Bibliographische Materialien zur Hochschulforschung.
Hochschulökonomie und Bildungsplanung.
Zweite erweiterte Auflage, 179 S.
Erschienen 1968.

9B Susanne Kleemann
Bibliographische Materialien zur Hochschulforschung.
Sozialisationsprozesse und Einstellungsveränderungen in der
Hochschule am Beispiel USA.
178 S. Erschienen 1969.

10 Klaus Herzog und Guy Oddie (OECD)
Technologische oder ökonomische Lösung des Schulbau-
problems.
Wirtschaftlichkeit im Schulbau.
307 S. Erschienen 1968.

11 Werner Kalb
Stiftungen und Bildungswesen in den USA.
246 S. Erschienen 1968.

12 Wolfgang Edelstein, Fritz Sang und Werner Stegelmann
Unterrichtsstoffe und ihre Verwendung in der 7. Klasse der
Gymnasien in der BRD (Teil I).
Eine empirische Untersuchung.
319 S. Erschienen 1968.

13 Klaus Huhse
Theorie und Praxis der Curriculum-Entwicklung.
Ein Bericht über Wege der Curriculum-Reform in den USA
mit Ausblicken auf Schweden und England.
227 S. Erschienen 1968.

14 Willi Voelmy
Systematische Inhaltsanalysen von Quellentexten zum Poly-
technischen Unterricht in der zehnklassigen allgemeinbil-
denden polytechnischen Oberschule der DDR 1959 bis 1966.
139 S. Erschienen 1968.

15 Hedwig Rudolph
Finanzierungsaspekte der Bildungsplanung dargestellt am Bei-
spiel des Schulsystems in Bayern.
146 S. Erschienen 1969.

16 Franz Scherer
Ökonomische Beiträge zur wissenschaftlichen Begründung der
Bildungspolitik.
193 S. Erschienen 1969.

17 Klaus Hüfner
Traditionelle Bildungsökonomie und systemorientierte
Bildungsplanung.
201 S. Erschienen 1969.

18 Ulrich Oevermann
Sprache und soziale Herkunft.
Ein Beitrag zur Analyse schichtenspezifischer Sozialisations-
prozesse und ihrer Bedeutung für den Schulerfolg.
327 S. Erschienen 1970 (übernommen in die edition
suhrkamp als Nr. 519).

19 Wolfgang Berger
Zur Theorie der Bildungsnachfrage.
Ein Beitrag zur Identifizierung der Determinanten privater
Nachfrage nach formaler Bildung.
162 S. Erschienen 1969.

20 Adolf Kell
Die Vorstellungen der Verbände zur Berufsausbildung
(2 Bände).
616 S. insgesamt. Erschienen 1970.

21 Frank Händle
Management in Forschung und Entwicklung. Bibliogra-
phische Materialien mit einer Einführung.
167 S. Erschienen 1971.

22 Peter Müller
Dokumentation zur Lehrerbildung (2 Bände).
550 S. insgesamt. Erschienen 1971.

23 Wolfgang Armbruster
Arbeitskräftebedarfsprognosen als Grundlage der Bildungs-
planung.
Eine kritische Analyse.
210 S. Erschienen 1971.

24 Hartmut J. Zeiher
Unterrichtsstoffe und ihre Verwendung in der 7. Klasse des
Gymnasiums in der BRD (Teil II).
Deutschunterricht.
261 S. Erschienen 1972.

25 Claus Oppelt, Gerd Schrick und Armin Bremmer
Gelernte Maschinenschlosser im industriellen Produktions-
prozeß.
Determinanten beruflicher Autonomie an Arbeitsplätzen von
Facharbeitern und Technischen Angestellten in der West-
berliner Industrie.
184 S. Erschienen 1972.

26 Annegret Harnischfeger
Die Veränderung politischer Einstellungen durch Unterricht.
Ein Experiment zur Beeinflussung der Nationbezogenheit.
268 S. Erschienen 1972.

27 Enno Schmitz
Das Problem der Ausbildungsfinanzierung in der neo-
klassischen Bildungsökonomie.
127 S. Erschienen 1973.

28 Doris Elbers
Curriculumreformen in den USA.
Ein Bericht über theoretische Ansätze und praktische Reform-
verfahren mit einer Dokumentation über Entwicklungs-
projekte.
205 S. Erschienen 1973.

29 Peter Matthias
Determinanten des beruflichen Einsatzes hochqualifizierter
Arbeitskräfte.
Zur Berufssituation von Diplom-Kaufleuten.
205 S. Erschienen 1973.

30 Jens Naumann
Medien-Märkte und Curriculumrevision in der BRD.
Eine bildungsökonomische Studie zu den Entstehungs-
bedingungen und Verbreitungsmechanismen von Lernmitteln
und Unterrichtstechnologien.
179 S. Erschienen 1974.

31 Gisela Klann
Aspekte und Probleme der linguistischen Analyse schichten-
spezifischen Sprachgebrauchs.
304 S. Erschienen 1975.

32 Dirk Hartung und Reinhard Nuthmann
Status- und Rekrutierungsprobleme als Folgen der Expan-
sion des Bildungssystems.
184 S. Erschienen 1975.

33 Helmut Köhler
Lehrer in der Bundesrepublik Deutschland.
Eine kritische Analyse statistischer Daten über das Lehr-
personal an allgemeinbildenden Schulen.
270 S. Erschienen 1975.

34A Hartmut-W. Frech
Empirische Untersuchungen zur Ausbildung von Studien-
referendaren.
Berufsvorbereitung und Fachsozialisation von Gymnasial-
lehrern.
298 S. Erschienen 1976.

34B Roland Reichwein
Empirische Untersuchungen zur Ausbildung von Studien-
referendaren.
Traditionelle und innovatorische Tendenzen in der beruf-
lichen Ausbildungsphase von Gymnasiallehrern.
352 S. Erschienen 1976.

34C Karl-Heinz Hebel
Empirische Untersuchungen zur Ausbildung von Studien-
referendaren.
Methodologische Implikationen einer Feldstudie zur
Gymnasiallehrerausbildung, konkretisiert an ausgewählten
Beispielen zur Berufsmotivation.
211 S. Erschienen 1976.

35 Hans-Ludwig Freese
Schulleistungsrelevante Merkmale der häuslichen
Erziehungsumwelt.
Ergebnisse einer empirischen Untersuchung über Jungen
und Mädchen der 7. Klasse des Gymnasiums.
141 S. Erschienen 1976.

36 Peter Siewert
Kostenrechnung für Schulen in öffentlicher Trägerschaft.
Fragen und Ansätze.
105 S. Erschienen 1976.

37 Claus Oppelt
Ingenieure im Beruf.
Eine empirische Analyse zertifikatsspezifischer Unter-
schiede im beruflichen Einsatz technischer Arbeitskräfte.
223 S. Erschienen 1976.

Neuere Bücher aus dem Max-Planck-Institut für Bildungsforschung*

I. Klett-Cotta Verlag, Stuttgart

Klaus Hüfner, Jens Naumann, Helmut Köhler und
Gottfried Pfeffer
Hochkonjunktur und Flaute: Bildungspolitik in der Bundes-
republik Deutschland 1967–1980.
361 S. Erschienen 1986.

Achim Leschinsky und Peter M. Roeder
Schule im historischen Prozeß – Zum Wechselverhältnis von
institutioneller Erziehung und gesellschaftlicher Entwicklung.
545 S. Erschienen 1976 (vergriffen; erhältlich ist noch die
Ullstein-Taschenbuch-Ausgabe Nr. 39055, erschienen 1983).

Knut Nevermann
Der Schulleiter.
Juristische und historische Aspekte zum Verhältnis von
Bürokratie und Pädagogik.
314 S. Erschienen 1982.

Gerd Sattler
Englischunterricht im FEGA-Modell.
Eine empirische Untersuchung über inhaltliche und methodische
Differenzierung an Gesamtschulen.
355 S. Erschienen 1981.

Diether Hopf
Mathematikunterricht.
Eine empirische Untersuchung zur Didaktik und Unterrichts-
methode in der 7. Klasse des Gymnasiums.
251 S. Erschienen 1980.

Christel Hopf, Knut Nevermann und Ingo Richter
Schulaufsicht und Schule.
Eine empirische Analyse der administrativen Bedingungen
schulischer Erziehung.
428 S. Erschienen 1980.

Max-Planck-Institut für Bildungsforschung, Projektgruppe
Bildungsbericht (Hrsg.)
Bildung in der Bundesrepublik Deutschland.
Daten und Analysen.
Bd. 1: Entwicklungen seit 1950.
Bd. 2: Gegenwärtige Probleme.
1404 S. Erschienen 1980.

Helga Zeiher, Hartmut J. Zeiher und Herbert Krüger
Textschreiben als produktives und kommunikatives Handeln.
Bd. I: Beurteilung von Schülertexten.
254 S. Erschienen 1979.

Helga Zeiher, Hartmut J. Zeiher und Herbert Krüger
Textschreiben als produktives und kommunikatives Handeln.
Bd. III: Synergetischer Textunterricht.
170 S. Erschienen 1979.

Klaus Hüfner und Jens Naumann
Konjunkturen der Bildungspolitik in der Bundesrepublik
Deutschland.
Bd. 1: Der Aufschwung (1960–1967).
307 S. Erschienen 1977.

Lothar Krappmann
Soziologische Dimensionen der Identität.
Strukturelle Bedingungen für die Teilnahme an Interaktions-
prozessen.
231 S. Erschienen 1971 (Konzepte der Humanwissenschaften,
6. Auflage 1982).

Peter Damerow, Ursula Elwitz, Christine Keitel
und Jürgen Zimmer
Elementarmathematik: Lernen für die Praxis?
Ein exemplarischer Versuch zur Bestimmung fachüber-
schreitender Curriculumziele.
182 und 47 S. Erschienen 1974.

II. Campus Verlag, Frankfurt/New York

Hans-Peter Blossfeld, Alfred Hamerle und Karl Ulrich Mayer
Ereignisanalyse.
Statistische Theorie und Anwendung in den Wirtschafts- und
Sozialwissenschaften.
290 S. Erschienen 1986.

Christel Hopf, Knut Nevermann und Ingrid Schmidt
Wie kamen die Nationalsozialisten an die Macht.
Eine empirische Analyse von Deutungen im Unterricht.
344 S. Erschienen 1985.

Hans-Peter Blossfeld
Bildungsexpansion und Berufschancen.
Empirische Analysen zur Lage der Berufsanfänger in der
Bundesrepublik.
191 S. Erschienen 1985.

III. Andere Verlage

Margret M. Baltes and Paul B. Baltes (Eds.)
The psychology of control and aging.
415 pp. N. J. Erlbaum, Hillsdale, N. J., 1986.

Paul B. Baltes, David L. Featherman and
Richard M. Lerner (Eds.)
Life-span development and behavior.
334 pp. Vol. 7. N. J. Erlbaum, Hillsdale, N. J., 1986.

Axel Funke, Dirk Hartung, Beate Krais und Reinhard Nuthmann
Karrieren außer der Reihe.
Bildungswege und Berufserfolg von Stipendiaten der gewerk-
schaftlichen Studienförderung.
256 S. Bund-Verlag GmbH, Köln 1986.

Jürgen Staupe
Parlamentsvorbehalt und Delegationsbefugnis.
Zur „Wesentlichkeitstheorie" und zur Reichweite legislativer
Regelungskompetenz, insbesondere im Schulrecht.
419 S. Duncker & Humblot, Berlin 1986.

Ernst-H. Hoff
Arbeit, Freizeit und Persönlichkeit.
Wissenschaftliche und alltägliche Vorstellungsmuster.
229 S. Verlag Hans Huber, Bern/Stuttgart/Toronto 1986.

John R. Nesselroade and Alexander von Eye (Eds.)
Individual development and social change: Explanatory analysis.
380 pp. Academic Press, New York 1985.

Ernst-H. Hoff, Lothar Lappe und Wolfgang Lempert (Hrsg.)
Arbeitsbiographie und Persönlichkeitsentwicklung.
288 S. Verlag Hans Huber, Bern/Stuttgart/Toronto 1985.

Michael Jenne
Music, Communication, Ideology.
185 pp. Birch Tree Group Ltd., Princeton, N. J., 1984.

Arbeitsgruppe am Max-Planck-Institut für Bildungsforschung
Das Bildungswesen in der Bundesrepublik Deutschland.
Ein Überblick für Eltern, Lehrer, Schüler.
312 S. Aktualisierte und erweiterte Neuausgabe.
Rowohlt Taschenbuch Verlag, Reinbek 1984.

Max Planck Institute for Human Development and Education
Between Elite and Mass Education.
Education in the Federal Republic of Germany.
348 pp. State University of New York Press, Albany 1983.

---

* Einschließlich der noch im Buchhandel erhältlichen Bände der „Veröffentlichungen des Max-Planck-Instituts für
  Bildungsforschung".

Margit Osterloh
Handlungsspielräume und Informationsverarbeitung.
369 S. Verlag Hans Huber, Bern/Stuttgart/Toronto 1983.

Bernhard Schmitz
Zeitreihenanalyse in der Psychologie.
Verfahren zur Veränderungsmessung und Prozeßdiagnostik.
304 S. Deutscher Studien Verlag, Beltz Verlag,
Weinheim/Basel 1987.

Hans-Uwe Hohner
Kontrollbewußtsein und berufliches Handeln.
201 S. Verlag Hans Huber, Bern/Stuttgart/Toronto 1987.

Frühere Buchreihe VERÖFFENTLICHUNGEN DES MAX-PLANCK-INSTITUTS FÜR BILDUNGSFORSCHUNG, zuvor: Texte und Dokumente zur Bildungsforschung.

Von diesen im Klett-Cotta Verlag, Stuttgart, erschienenen — und inzwischen vergriffenen — Büchern sind teilweise noch Restexemplare im Institut erhältlich.

Günter Palm
Die Kaufkraft der Bildungsausgaben.
Ein Beitrag zur Analyse der öffentlichen Ausgaben für Schulen und Hochschulen in der Bundesrepublik Deutschland 1950 bis 1962.
183 S. Erschienen 1966.

Torsten Husén und Gunnar Boalt
Bildungsforschung und Schulreform in Schweden.
254 S. Erschienen 1968.

James B. Conant
Bildungspolitik im föderalistischen Staat — Beispiel USA.
130 S. Erschienen 1968.

Henry Chauncey und John E. Dobbin
Der Test im modernen Bildungswesen.
176 S. Erschienen 1968.

Michael Jenne, Marlis Krüger und Urs Müller-Plantenberg
Student im Studium.
Untersuchungen über Germanistik, Klassische Philologie und Physik an drei Universitäten.
Mit einer Einführung von Dietrich Goldschmidt.
464 S. Erschienen 1969.

Ulrich K. Preuß
Zum staatsrechtlichen Begriff des Öffentlichen untersucht am Beispiel des verfassungsrechtlichen Status kultureller Organisationen.
229 S. Erschienen 1969.

Ingo Richter
Die Rechtsprechung zur Berufsausbildung.
Analyse und Entscheidungssammlung.
623 S. Erschienen 1969.

Klaus Hüfner und Jens Naumann (Hrsg.)
Bildungsökonomie — Eine Zwischenbilanz.
Economics of Education in Transition.
Friedrich Edding zum 60. Geburtstag.
275 S. Erschienen 1969.

Helge Lenné
Analyse der Mathematikdidaktik in Deutschland.
Aus dem Nachlaß hrsg. von Walter Jung in Verbindung mit der Arbeitsgruppe für Curriculum-Studien.
446 S. Erschienen 1969.

Wolfgang Dietrich Winterhager
Kosten und Finanzierung der beruflichen Bildung.
161 S. Erschienen 1969.

Philip H. Coombs
Die Weltbildungskrise.
248 S. Erschienen 1969.

Klaus Hüfner (Hrsg.)
Bildungsinvestitionen und Wirtschaftswachstum.
Ausgewählte Beiträge zur Bildungsökonomie.
356 S. Erschienen 1970.

Jens Naumann (Hrsg.)
Forschungsökonomie und Forschungspolitik.
Ausgewählte amerikanische Beiträge.
482 S. Erschienen 1970.

Matthias Wentzel
Autonomes Berufsausbildungsrecht und Grundgesetz.
Zur Rechtsetzung der Industrie- und Handelskammern und Handwerksorganisationen in der Bundesrepublik.
229 S. Erschienen 1970.

Dieter Berstecher
Zur Theorie und Technik des internationalen Vergleichs.
Das Beispiel der Bildungsforschung.
123 S. Erschienen 1970.

Bernhard Dieckmann
Zur Strategie des systematischen internationalen Vergleichs.
Probleme der Datenbasis und der Entwicklungsbegriffe.
188 S. Erschienen 1970.

Dirk Hartung, Reinhard Nuthmann und Wolfgang Dietrich Winterhager
Politologen im Beruf.
Zur Aufnahme und Durchsetzung neuer Qualifikationen im Beschäftigungssystem.
250 S. Erschienen 1970.

Saul B. Robinsohn u. a.
Schulreform im gesellschaftlichen Prozeß.
Ein interkultureller Vergleich.
Bd. I: Bundesrepublik, DDR, UdSSR.
602 S. insgesamt. Erschienen 1970.

Saul B. Robinsohn u. a.
Schulreform im gesellschaftlichen Prozeß.
Ein interkultureller Vergleich.
Bd. II: England und Wales, Frankreich, Österreich, Schweden.
595 S. Erschienen 1975.

Klaus Hüfner und Jens Naumann (Hrsg.)
Bildungsplanung: Ansätze, Modelle, Probleme.
Ausgewählte Beiträge.
362 S. Erschienen 1971.

Pierre Bourdieu und Jean-Claude Passeron
Die Illusion der Chancengleichheit.
Untersuchungen zur Soziologie des Bildungswesens am Beispiel
Frankreichs.
302 S. Erschienen 1971.

Wolfgang Karcher
Studenten an privaten Hochschulen.
Zum Verfassungsrecht der USA.
240 S. Erschienen 1971.

Marianne von Rundstedt
Studienförderung.
Ein Vergleich der Förderungssysteme und Leistungen in der
Bundesrepublik Deutschland, Belgien, Frankreich, England und
Wales und in den Niederlanden.
189 S. Erschienen 1971.

Helga Zeiher
Gymnasiallehrer und Reformen.
Eine empirische Untersuchung über Einstellungen zu Schule und
Unterricht.
279 S. Erschienen 1973.

Ingo Richter
Bildungsverfassungsrecht.
Studien zum Verfassungswandel im Bildungswesen.
365 S. Erschienen 1973.

Wolfgang Lempert und Wilke Thomssen
Berufliche Erfahrung und gesellschaftliches Bewußtsein.
Untersuchungen über berufliche Werdegänge, soziale Einstel-
lungen, Sozialisationsbedingungen und Persönlichkeitsmerkmale
ehemaliger Industrielehrlinge (Bd. I).
521 S. Erschienen 1974.

Detlef Oesterreich
Autoritarismus und Autonomie.
Untersuchungen über berufliche Werdegänge, soziale Einstel-
lungen, Sozialisationsbedingungen und Persönlichkeitsmerkmale
ehemaliger Industrielehrlinge (Bd. II).
248 S. Erschienen 1974.

Jürgen Raschert
Gesamtschule: ein gesellschaftliches Experiment.
Möglichkeiten einer rationalen Begründung bildungspolitischer
Entscheidungen durch Schulversuche.
217 S. Erschienen 1974.

Ulrich Teichler
Geschichte und Struktur des japanischen Hochschulwesens
(Hochschule und Gesellschaft in Japan, Bd. I).
385 S. Erschienen 1975.

Ulrich Teichler
Das Dilemma der modernen Bildungsgesellschaft.
Japans Hochschulen unter den Zwängen der Statuszuteilung
(Hochschule und Gesellschaft in Japan, Bd. II).
483 S. Erschienen 1976.

Michael Jenne
Musik — Kommunikation — Ideologie.
Ein Beitrag zur Kritik der Musikpädagogik.
202 S. Erschienen 1977.

Fritz Sang
Elternreaktionen und Schulleistung.
Bedingungen und Konsequenzen Leistung erklärender
Attributionen.
158 S. Erschienen 1977.

Peter Damerow
Die Reform des Mathematikunterrichts in der Sekundarstufe I.
Eine Fallstudie zum Einfluß gesellschaftlicher Rahmen-
bedingungen auf den Prozeß der Curriculumreform.
Bd. I: Reformziele, Reform der Lehrpläne.
368 S. Erschienen 1977.

Hartmut-W. Frech und Roland Reichwein
Der vergessene Teil der Lehrerbildung.
Institutionelle Bedingungen und inhaltliche Tendenzen im
Referendariat der Gymnasiallehrer.
326 S. Erschienen 1977.

Enno Schmitz
Leistung und Loyalität.
Berufliche Weiterbildung und Personalpolitik in Industrie-
unternehmen.
278 S. Erschienen 1978.

Jürgen Baumert und Jürgen Raschert
in Zusammenarbeit mit Diether Hopf, Jens Naumann
und Helga Thomas
Vom Experiment zur Regelschule.
Schulplanung, Curriculumentwicklung der Lehrerfortbildung
in Zusammenarbeit von Lehrern und Verwaltung bei der
Expansion der Berliner Gesamtschule.
276 S. Erschienen 1978.

Im Institut erhältlich sind noch Restexemplare der Taschenbuch-
Ausgabe des „1. Bildungsberichts":

Max-Planck-Institut für Bildungsforschung, Projektgruppe
Bildungsbericht (Hrsg.)
Bildung in der Bundesrepublik Deutschland.
Daten und Analysen.
Bd. 1: Entwicklungen seit 1950.
Bd. 2: Gegenwärtige Probleme.
1404 S. Erschienen 1980 (Rowohlt Taschenbuch Nr. 7337
und Nr. 7338, vergriffen).