How reliable is pooled analysis in political economy? The globalization-welfare state nexus revisited

BERNHARD KITTEL[1] & HANNES WINNER[2]

[1]*University of Amsterdam, The Netherlands;* [2]*University of Innsbruck, Austria*

Abstract. This article analyzes various pitfalls that arise in the application of panel data methods in comparative political economy. Empirically, we refer to the debate on the globalization-welfare state nexus by re-assessing a study by Garrett and Mitchell ('Globalization, Government Spending and Taxation in the OECD', *European Journal of Political Research* 39(1) (2001): 145–177). We discuss the properties of specifications with time invariate political variables, dynamic models with nonstationary data, and autocorrelated residuals. We demonstrate that the findings of previous empirical studies are often driven by mis-specifications. Presenting a statistically well-behaved model, we find evidence that government spending is primarily driven by the state of the domestic economy. Neither partisan effects nor the international economic environment have affected public expenditure considerably.

Introduction

During the 1990s, advances in the statistical methodologies used for assessing hypotheses on the interrelations between economic development and political institutions have been impressive. The availability of data for a set of usually about 15 to 25 OECD countries over a period of up to about 40 years has placed techniques of panel data analysis in the forefront of applied research. It is no exaggeration to say that it has become difficult to defend not using panel data in the analysis of comparative political economy.

At least, the importance of panel data analysis is well-documented in recent studies on the relationship between globalization and the several dimensions of the welfare state. There are two conflicting views on this issue (Schulze & Ursprung 1999). The *efficiency hypothesis* states that the emerging internationalization of the economies induces a downward pressure on tax rates and government spending (e.g., Bretschger & Hettich 2002; Rodrik 1997, 1998; Swank & Steinmo 2002). In contrast, the *compensation hypothesis* (see Cameron 1978; Katzenstein 1985) claims that increasing factor mobility is associated with a higher demand on social security, which in turn causes an upward shift of taxation and spending levels (see Garrett 1995, 1998; Hicks &

Published by Blackwell Publishing Ltd., 9600 Garsington Road, Oxford, OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

Swank 1992; Huber & Stephens 2001; Quinn 1997; Swank 1998; ambivalent evidence is given by Garrett and Mitchell 2001).

This article is motivated out of the concern that the use of panel data analysis as a universal remedy for all the problems of cross-country comparative analysis has led the field into an impasse. In the debate on the globalization-welfare state nexus, the impasse appears at three levels. First, little effort is made to develop a concise theory as to why specific variables are entered into a regression and how they are related to each other. Instead, we observe the discussion of singular, often ad hoc, propositions under the guise of hypotheses. Sometimes, this is associated with an unreflected use of theoretical concepts (for a critique, see Moses 2001). Second, we observe empirical specifications that rarely discuss the assumptions about the spherical nature of the data. Specifically, numerous studies simply take for granted the finding of Beck and Katz (1995) that panel data tend to violate the classical assumptions of the error term (groupwise heteroskedasticity, cross-sectional and serial correlation) and regard autoregressive models with panel-corrected standard errors as a panacea. However, Beck and Katz (1995: 645; see also Greene 2000: 592–607) argue that it is necessary to test the assumptions on the error term *before* applying such a correction – a suggestion that is often ignored in applied empirical work. And third, the problems induced by the time dimension of pooled data, notably due to nonstationary data, are often ignored. Consequently, the combination of weak theoretical reasoning and an ambivalent statistical foundation results in highly problematic conclusions.

This article deals *exclusively* with the second- and third-level, the *statistical modeling* of the globalization-welfare state nexus. Although we suspect that this is close to being a paramount problem in quantitative contributions to comparative political economy, we will not address the topic in a generalized manner. Instead, we discuss the problem by re-analyzing a data set used in a paper by Geoffrey Garrett and Deborah Mitchell (2001). Utilizing panel data for 17 OECD countries and the time period 1961 to 1993, they find evidence that globalization and the partisan composition of governments have a systematic impact on revenue and expenditure levels. However, this study – though stimulating in many respects – clearly exemplifies the various pitfalls we may encounter when analyzing panel data. Hence, we use this paper as a reference for a discussion of the issues we believe to be important in the course of an analysis of panel data in comparative political economy.

Our emphasis on model selection, specification and estimation issues implies that we do not directly address the question concerning which substantive variables should be included to explain public expenditure figures (see, e.g., Mueller 2003: Chapter 21). We do not discuss the theoretical justification of the inclusion of particular variables, but proceed from the variables included by Garrett and Mitchell (2001). Our contribution is to explore the statistical properties of different assumptions salient to panel data. Reassessing the empirical results of Garrett and Mitchell, we demonstrate that model specifications widely used in the globalization literature typically suffer from mis-specifications. Worse, the parameter estimates are usually biased, inefficient and inconsistent, thereby placing the validity of the inferences drawn from these studies at stake.

This article is organized as follows. The next section introduces a static panel data model and discusses the implications of controlling for cross-sectional and time-specific idiosyncrasies. The third section re-estimates the dynamic specification of Garrett and Mitchell (2001) and demonstrates the major shortcomings therein. The fourth section shifts to a dynamic model in first differences and discusses the consequences of this approach. The fifth section addresses issues of the country- and time-dependency of our findings. The final section presents our conclusions.

The nature of the problem: (Almost) time invariant variables

We focus our discussion on one of the models presented by Garrett and Mitchell. More specifically, we chose the 'total spending' model in the first column of their Table 5 (Garrett & Mitchell 2001: 166), reprinted as the column indicated by 'G/M' in our Table 2, on which they base their discussion of the relationship between total government expenditure (defined as a percentage of GDP) and the partisan composition of government as well as economic internationalization. Partisan composition is measured as the share of leftist and Christian Democratic parties in government; globalization is proxied by trade openness (imports plus exports related to GDP), the share of imports from low-wage countries on total imports and flows of foreign direct investment (FDI). In addition, Garrett and Mitchell add three control variables: the unemployment rate, GDP growth per capita and the dependency ratio (defined as the share of citizens aged above 60 and below 19). Although the choice of the total spending model seems somewhat arbitrary given the other models presented in their paper, two factors make it a reasonable choice. First, total government expenditure is a crucial dependent variable in the substantive argument made by Garrett and Mitchell (2001); and second, the model has properties along which the problems of model specification in the pooled context can be exemplified rather clearly.[1]

First, we follow Garrett and Mitchell and assume poolability of the data in the sense that the estimation of one coefficient over 'space' and 'time' for each variable is a reasonable approximation to reality (see Hsiao 2003: 14–15).[2] In the next step, we impose common slopes, but allow for varying intercepts (country and/or time effects). This represents the simplest type of panel data models, known as the 'fixed effects' or 'within' estimator. Country effects capture the unobserved country-specific variation in a country-specific intercept. Since this removes the average country effect, such a model focuses on the within-country variation over time, and the coefficients represent a cross-country average of the longitudinal effect. Time effects, in contrast, capture developments over time that are common to all countries. This 'within-time' estimator eliminates any common trends and external shocks to which all countries are jointly exposed. The coefficient refers to the average cross-sectional effect over time and takes into account shifts over time in the position of the countries relative to each other. The variance of a coefficient estimate becomes smaller, the greater the persistency of the relative position of the countries becomes on both the dependent and the independent variable. Combining both country and time intercepts in a single specification results in a model from which all unobserved country- and time-specific effects are removed. What remains is a 'pure' effect, stripped of all exogenous noise and all variance components that are constant in either the time or cross-section dimension. Finally, like the within-time estimator, the 'between estimator' focuses on the cross-sectional dimension, but first removes the within-country variation by averaging over time and then performs a cross-sectional regression on the country means.

Table 1 presents the results of the within and between estimator and compares them to the pooled OLS ones. Here, we apply a static model in levels and test whether the structure of the error term is captured adequately. Since panel data can exhibit serial correlation, cross-sectional correlation and groupwise heteroskedasticity (Greene 2000: 592–608), we expect to find such a structure in the residuals. The lower block of Table 1 reports the corresponding specification tests. They suggest that we have to include both country and time effects, and that the residuals do indeed reveal the expected panel structure.[3] Hence, the parameter estimates in the first two columns of Table 1 are no longer BLUE, and also the test statistic turns out to be invalid. Nevertheless, it is worthwhile to explore the coefficient estimates because they reveal a core problem of pooled analysis with political and institutional variables, which tend to be constant or not varying considerably over time. This is the case for leftist and Christian Democratic cabinet portfolios. If we ignore the bias in the standard errors, both coefficients are positive and rather high in the pooled specification. However, including country and time effects changes the sign for Leftist parties and makes the party coefficient for Christian Democratic parties insignificant. Similar conclusions have to be drawn for the dependency ratio and low-wage imports. In contrast, the coefficients of unemployment,

Table 1. Static specification in levels (dependent variable: total government expenditure as percentage of GDP)

	Pooled OLS	Fixed effects FE(CT)	Between estimator	
			(1)	(2)
Unemployment	0.932*** (0.093)	0.706*** (0.080)	1.673 (1.100)	1.770** (0.678)
GDP/capita growth	–0.901*** (0.130)	–0.510*** (0.069)	0.978 (2.711)	–
Dependency ratio	–0.445*** (0.124)	1.329*** (0.111)	–0.355 (1.346)	–
Left cabinet portfolios	5.802*** (0.872)	–1.261*** (0.459)	23.445** (9.651)	25.484*** (7.078)
Christian Democrat portfolios	2.876** (1.318)	–0.784 (1.049)	5.362 (10.080)	11.378* (5.992)
Trade	0.129*** (0.015)	0.025 (0.020)	0.083 (0.115)	–
Low wage imports	–0.157*** (0.048)	0.338*** (0.035)	–0.029 (0.382)	–
Foreign direct investment	0.093 (0.216)	0.142 (0.131)	0.915 (2.478)	–
R^2	0.584	0.941	0.665	0.573
N (observations)	529	529	17	17
F(C)	–	146.66***	–	–
F(T)	–	23.95***	–	–
F(CT)	–	60.76***	–	–
Cross-sectional correlation[a]	537.09***	487.03***	–	–
Groupewise heteroskedasticity[b]	1745.20***	495.12***	–	–
Autocorrelation (AR1)[c]	471.48***	365.72***	–	–
Heteroskedasticity[d]	–	–	–	6.40

Notes: FE(CT) = fixed country and time effects; constant and fixed effects not reported; * $\alpha \leq 0.10$; ** $\alpha \leq 0.05$; *** $\alpha \leq 0.01$; F(C) = F test for country effects; F(T) = F test for time effects; F(CT) = F test for country and time effects; [a] χ^2 (136), Breusch-Pagan LM, see Greene (2000: 601); [b] χ^2 (17), modified Wald test, see Greene (2000: 598); [c] χ^2 (1), LM test for AR1, see Baltagi (2001: 95); [d] χ^2 (9), White (1980) test for heteroskedasticity, leftist and Christian Democratic cabinet portfolios expressed as ratios.

economic growth per capita and trade are much more stable. The effect of FDI remains insignificant.

Here, we only discuss the party coefficients because they are the core variables in most of the globalization literature. Their sign and significance is decisive in the substantive interpretation of the salience of partisan effects that Garrett (1998) strongly advocates. Why is the sign of partisanship unstable? The answer is twofold. First, in some of the countries – Canada and the United States for leftist cabinet portfolios; and Australia, Canada, Ireland, Japan and the United Kingdom for Christian Democratic portfolios – these variables do not vary over time. This implies that these countries do not enter the partisan estimates in a specification with fixed country effects, which remove all cross-sectional variation. Since Canada and the United States both combine no leftist government portfolios with low levels of government expenditure, their exclusion from the sample in practice may change the estimate of the partisan effect. In addition, the partisanship variables have extremely small variation in a couple of other countries (e.g., Denmark, Finland, Switzerland). Second, by focusing on the longitudinal dimension, the static model in levels assumes an immediate shift of government expenditure each time the partisan composition of the government changes. While this is somewhat implausible, notably due to the persistency of public expenditure, it is not surprising that the significance of our estimates is rather low. Since the time effects focus on the cross-sectional dimension, the differences in total government expenditure between the countries are captured without interference of the general upward trend (which increases the variance of the coefficient estimates in pooled OLS), leading to more pronounced effects. The FE(CT) estimates combine both the within-time and the within-country variation, where the weight depends on the relative variation in the two dimensions (see Baltagi 2001: 31–33). Hence, from a substantive perspective, the size and the sign of the estimates in the FE(CT) specification depend on the practical exclusion of particular countries.[4] We will come back to this issue when discussing the robustness of the dynamic specification.

At this point, it is useful to look at the results obtained from the between estimator. Since this is a pure, cross-sectional regression on the country means, we do not have to bother about the panel structure. The last two columns of Table 1 report the results of the between regressions. Despite the low degrees of freedom, we find a highly significant 23 percentage point difference between 0 per cent and 100 per cent leftist portfolios. The model captures 67 per cent of the variation in the dependent variable. Since most coefficients are far from being significant, we re-estimate a restricted specification including only the partisan effects and unemployment, which failed the 10 per cent significance level only by a narrow margin. In the restricted specification (last column of

Table 1), the coefficient of Christian Democratic portfolios becomes highly significant, too. Since the White (1980) test does not suggest the presence of heteroskedasticity, we stick with this specification.

In contrast to the between estimator, the FE model cannot clearly confirm the presence of partisan effects. Garrett and Mitchell (2001: 169) decide to interpret the country intercepts as an 'equilibrium' level of spending. This may make sense if a substantive meaning can be attributed to the intercepts – for instance, if they can be read as long-term, steady-state solutions (see, e.g., Daveri & Tabellini 2000: 72 for such an interpretation in the context of unemployment). In the present case, such an interpretation seems to be inconsistent because Garrett and Mitchell do not provide a theoretical model from which a steady state of government expenditure could be inferred. In a less strict line of reasoning, one might claim that the country intercepts simply produce the average of the dependent variable as it is conditional on the other explanatory variables. However, how do we deal with changes in government composition when interpreting the intercepts? Since the fitted country average is the net of the effect of all regressors, the average level cannot be attributed to partisan effects because these are already included in the model.[5]

Overall, the static model reveals a fundamental problem with regard to the relationship between partisanship and public spending. While the between estimator indicates the presence of partisan effects on government expenditure, the fixed effects model does not confirm this finding. A standard solution is to instrument the time invariant partisanship variables with the remaining independent ones (Hausman & Taylor 1981). However, since we are interested in the cross-sectional variation – as is typically the focus of comparative political economy – this solution is not viable here. An alternative would be to ignore the fixed country effects and to attempt to model the cross-sectional dimension via institutional variables that capture enough of the cross-sectional variation to make the inclusion of country effects insignificant. Unfortunately, such variables are usually not available.

A further complication: Dynamic model with autocorrelation and nonstationarities

To disentangle the short- and long-run effects of globalization and partisan composition on the extent of the welfare state requires one to specify a dynamic model (i.e., to include a lagged dependent variable). This, in turn, raises additional difficulties, especially in the presence of autocorrelation. Basically, there are two ways to deal with this issue (Beck & Katz 1995, 1996). First, autocorrelation can be regarded as a nuisance in the residuals that has to be corrected. Second, autocorrelation may indicate persistency in the dependent variable that can be captured by modelling an autoregressive process including a lagged dependent variable (see Kittel & Winner 2002, for a detailed discussion on the relationship between these solutions).

The results of both approaches are reported in Table 2. First, the autocorrelation coefficient ρ is quite large, 0.8 (see first column), and, compared to the FE(CT) specification of Table 1, the coefficients are shuffled again, notably those of the partisanship variables. The second column presents the findings of the second approach, the inclusion of a lagged dependent variable (dynamic specification). As in the static model, the F-statistic suggests choosing a model with country and time effects (FE(CT)). This is the specification that Garrett and Mitchell (2001: 166) presented in the first column of their Table 5 (and we reproduce in the second column of our Table 2),[6] except for the calculation of the standard errors that we will discuss later.

In the dynamic model, the results for partisan composition may be interpreted as follows. The short-run impact of a shift from 0 per cent to 100 per cent leftist portfolios in government is to reduce government expenditure by (insignificant) 0.23 percentage points. In the long run, this shift amounts to a decrease of 0.225/(1 – 0.914) = 2.62 percentage points. However, both effects are insignificant, which, in turn, confirm the results of the static FE-models (Table 1 and Column 1 of Table 2). Further, the test for groupwise heteroskedasticity is highly significant. Hence, it seems reasonable to follow the recommendation by Beck and Katz (1995: 638) to apply panel corrected standard errors. In Table 2, we reproduce the FE(CT) model using weighted least squares (the first step of the Parks-Kmenta FGLS procedure) as well as OLS with panel-corrected standard errors (PCSE). The difference between the standard errors of FE(CT) and the PCSE is minor. In effect, the PCSEs are somewhat more permissive than the FE(CT) standard errors and somewhat less optimistic than the WLS ones. This is what we would expect given that the PCSEs make use of the panel structure. While this makes them less efficient, it also makes them – as emphasized by Beck and Katz (1995) – more realistic than WLS.

We contend, however, that heteroskedasticity and cross-sectional correlation are not the real problems in these data. Beck and Katz (1995) stress that autocorrelation has to be handled adequately *before* their standard errors are calculated. In the models presented in Table 2, two problems are apparent. According to the test for serial correlation tabulated in Table 2, none of the specifications is free of autocorrelation. An auxiliary regression of the residuals on the lagged residuals still gives a highly significant autocorrelation of 0.148 and Durbin's m-test (Kmenta 1990: 333) gives a coefficient of 0.172, which is also highly significant. A lagged dependent variable in combination

Table 2. Dynamic specification in levels (dependent variable: total government expenditure as percentage of GDP)

	FE(CT)/PW	FE(CT)	G/M	WLS	PCSE
Lagged government expenditure	–	0.914*** (0.019)	0.911***	0.902*** (0.017)	0.914*** (0.024)
Unemployment	0.610*** (0.071)	0.021 (0.036)	0.033	0.061** (0.031)	0.021 (0.039)
GDP/capita growth	–0.196*** (0.028)	–0.365*** (0.028)	–0.362***	–0.367*** (0.025)	–0.365*** (0.032)
Dependency ratio	0.930*** (0.148)	0.084 (0.053)	0.097**	0.092** (0.040)	0.084* (0.045)
Left cabinet portfolios[a]	–0.199 (0.318)	–0.225 (0.190)	–0.003	–0.133 (0.172)	–0.225 (0.204)
Christian Democrat portfolios[a]	0.302 (0.900)	–0.416 (0.432)	–0.005	–0.534* (0.319)	–0.416 (0.373)
Trade	–0.021 (0.020)	–0.026** (0.008)	–0.030**	–0.029*** (0.008)	–0.026** (0.011)
Low wage imports	0.107*** (0.034)	0.027* (0.016)	0.023	0.036*** (0.013)	0.027 (0.017)
Foreign direct investment	–0.098 (0.079)	–0.007 (0.054)	0.006	–0.013 (0.048)	–0.007 (0.057)
R^2	–	0.990	–	0.999	0.999
N (observations)	529	529	529	529	529
F (Country effects)	714.27***	4.81***	–	–	–
F (Time effects)	498.45***	3.01***	–	–	–
F (Country and time effects)	1088.18***	4.28***	–	–	–
Cross-sectional correlation	771.76***	146.87	–	–	–
Groupwise heteroskedasticity	1001.69***	372.04***	–	–	–
Autocorrelation (AR1)	447.60***	12.01***	–	12.94***	12.01***
Rho	0.800	–	–	–	–

Notes: see Table 1; PW = Prais-Winsten transformed FE estimates; WLS = Weighted Least Squares; PCSE = Panel-corrected Standard Errors; G/M = Parameter estimates reported by Garrett and Mitchell (2001: 166); constant and fixed effects not reported; Rho = autocorrelation parameter. [a] The estimates and those reported by Garrett and Mitchell differ by a factor of 100 because we have redefined their percentage values as ratios in order to improve readability.

with autocorrelation leads to a situation where OLS is no longer unbiased and consistent (Greene 2000: 534; for applications see Achen 2000). In addition, inserting a lagged dependent variable in a model with fixed country effects induces an additional bias via the correlation between the lagged dependent variable and the individual effects (see Baltagi 2001: 129–130; Kiviet 1995; Nickell 1981). The severity of the 'Nickell' bias depends on the sample size (cross-sections and time periods) and on the magnitude of the autoregressive coefficient (for Monte Carlo evidence, see Judson & Owen 1999).[7] The coefficient of the lagged dependent variable is above 0.9. This implies that we are confronted with a specification that is riddled with bias in *all* parameter estimates. Hence, the substantive interpretation of the coefficients and their standard errors is, as such, meaningless. These considerations give strong evidence that we have to reject the autoregressive specifications of Table 2 and, by implication, that of Garrett and Mitchell (2001).

A further problem arises from the extremely high autoregressive coefficient, suggesting that we might well be faced with nonstationarities.[8] Table 3 presents the results of Levin-Lin (LL) and Im-Pesaran-Shin (IPS) tests for unit roots in the dependent variable (for details, especially on the difference between these tests, see Maddala & Wu 1999; see also Baltagi 2001: 236–239).[9] The upper panel refers to the levels, while the lower presents the first differences. According to LL, the data appear nonstationary. The IPS tests clearly confirm this finding. Demeaning (i.e., removing the time effects) does not substantively alter the conclusions. Although these tests are notorious for their low power (see Maddala & Kim 1998: 137), we tend to be cautious and assume that the expenditure data are nonstationary. If this is valid, we are faced with an additional source of bias and of potentially spurious relationships.

There are two approaches for dealing with nonstationary data. First, we could remain in the single equation case and take first differences in order to proceed with a dynamic specification in differences. Second, we could explore the possibility of cointegrating relationships. This, however, requires more theoretical elaboration of the expected long-run association between the potentially cointegrating variables.[10] Since the literature on panel cointegration is still in its early stages of development (e.g., Pedroni 2000; Kao & Chiang 2000), we leave this task to a future endeavour and focus on the shift to a model in first differences.[11]

A statistically viable solution: A dynamic specification in first differences

A model in first differences focuses on systematic associations between the annual changes in the variables (i.e., the short-term effects) while removing

Table 3. Panel unit roots tests

Total government expenditure: levels				
Levin & Lin	*coeff*	*t*	*t**	*p*
constant	–0.113	–6.104	–0.871	0.192
constant, trend	–0.170	–5.835	3.184	0.999
Im, Pesaran & Shin	*t-bar*	*cv10*	ψ	*p*
demeaned, no trend	–1.606	–1.780	–0.383	0.351
demeaned, trend	–1.589	–2.410	2.727	0.997
not demeaned, no trend	–1.226	–1.780	1.287	0.901
not demeaned, trend	–1.733	–2.410	2.056	0.980
Total government expenditure: first differences				
Levin & Lin	*coeff*	*t*	*t**	*p*
constant	–0.836	–15.201	–10.030	0.000
constant, trend	–0.911	–16.569	–9.201	0.000
Im, Pesaran & Shin	*t-bar*	*cv10*	ψ	*p*
demeaned, no trend	–3.684	–1.780	–9.498	0.000
demeaned, trend	–3.921	–2.410	–8.162	0.000
not demeaned, no trend	–3.580	–1.780	–9.044	0.000
not demeaned, trend	–3.797	–2.410	–7.584	0.000

Notes: Levin & Lin (1992) (LL) tests augmented by 1 lag, H0 = nonstationarity; coeff = coefficient on lagged levels; t = t-value of coeff; t* = transformed t-value ~N(0,1); p = p-value of t*. Im-Pesaran-Shin (2003) (IPS) tests augmented by 1 lag, H0 = nonstationarity; t-bar = mean of country-specific Dickey-Fuller tests; cv10 = 10 per cent critical value of IPS test; ψ = transformed t-bar statistic, ~N(0,1); p = p-value of ψ. These tests are performed on a restricted sample to obtain a balanced panel. The data for Switzerland over the years 1991 to 1993 are dropped. Dropping Switzerland and Norway, but retaining the years 1991 to 1993 did not change the results significantly, except for the not-demeaned IPS tests for the levels, for which the p-value of ψ became clearly significant if the trend is excluded, but clearly insignificant if the trend is included. These considerable changes due to small changes to the sample suggest that the power of these tests seems to be rather low.

all level variation (i.e., the long-run effect). The decision to move to this specification is *solely* motivated by the statistical properties of the underlying data. Note that our data are I(1), which justifies the choice of lag length 1 and, therefore, of first differencing.

First differencing the data crucially changes the interpretation of our estimation results, in particular for that on partisan composition. If we take the first difference of these variables, we lose the information about the strength of the left or Christian Democracy in government and keep only the information about its change. This implies that we make inferences about the effect that the size of the change of, say, leftist party strength in government has on the size of the change in public expenditure. By implication, this approach assumes that a reduction in leftist party strength *directly* translates into a decline in public expenditure. Given our knowledge from qualitative research about the persistency and path-dependency of government programmes (see, e.g., Pierson 1996), such an assumption is only plausible if we claim that governments can fully and immediately change the level of public expenditure at their will. Instead, it is more reasonable to assume that parties in government can influence not the level, but the growth, of public spending during their term in office. Thus, we are faced with a serious problem in the initial-level specification chosen by Garrett and Mitchell (2001), highlighted by the autoregression coefficient of near unity and the apparent nonstationarity of public expenditure: The level of public expenditure should not be related to the strength of leftist or Christian Democratic parties in government during their term in office, but to the *cumulative* share of these parties in past terms, as advocated by Huber and Stephens (2001). This variable, however, is nonstationary by definition in the sense of an ever-increasing and thus non-constant mean, which returns us to the problems discussed above. Yet the share of these parties during a term is, also by definition, the first difference of this cumulative measure. Therefore, by shifting to this perspective, we move to a more plausible expectation about the potential impact of partisan composition on public expenditure.

It remains an empirical question whether first differencing fully removes the cross-country and time variation that is captured by the *fixed country and time effects*. On the one hand, if there are systematic cross-country differences in the annual expenditure growth, fixed country effects should capture enough variation to attain statistical significance in an F-test. On the other hand, if there are common developments in public expenditure, we would expect significant time effects. The latter may well be the case in OECD countries: while the 1960s witnessed a massive increase of the public sector in virtually all OECD countries, the 1980s were characterized by attempts at bringing that increase to a halt. Such policy reorientations – though to different degrees and with some variation in the exact timing for each country – are captured by time effects. Hence, we still test for fixed effects.

The results of the first difference specifications are presented in Table 4. The first two columns report the estimates of the fixed effects model. A striking feature of the first difference model is, however, that the estimates do not vary considerably, suggesting a robust empirical relationship. In all specifications, we find a highly significant impact of our core economic variables. Changes in unemployment are positively associated with changes in government expenditure, while the opposite holds for economic growth. For example, in the specification with country and time effects (first column), a change of the economic growth rate (per capita) by 1 percentage point is associated with a decrease in the change in government spending by 0.21 percentage points. Substantively, this indicates an income elasticity for public goods of less than unity, a result that is in line with most empirical findings (see Mueller 2003: 507). The change in the dependency ratio turns out to be positively correlated with government growth, as expected, and statistically significant. In contrast, apart from FDI, the political as well as the globalization variables are of limited importance for explaining government spending. Further, as shown in the F-tests on the significance of the country and time effects, we have to consider time effects in our model. Since the F-test on the country effects in the FE(CT) specification indicates that the country effects are insignificant – which is not unexpected since taking first differences wipes out the cross-country effects – we end up with a model including time effects only. Apart from the shift to the difference specification, which is necessary as pointed out previously, this constitutes a second major difference to the specification of Garrett and Mitchell (2001).

The last two columns of Table 4 present the estimates of the fixed time effects model, taking heteroskedasticity and cross-sectional correlation into account. First, we estimated the WLS-transformed model, then we used the PCSEs as proposed by Beck and Katz (1995). The WLS standard errors are much lower than those of the PCSE. This finding is in line with Beck and Katz, who have emphasized that WLS gives downward-biased standard errors. As outlined above, Beck and Katz have also stressed that in a correctly specified model, the standard errors of the OLS specification should not deviate considerably from those of the PCSE. This is exactly the case in our situation, which, in turn, is an indication that our model in Table 4 is correctly specified.

To conclude, in the first difference specification, we find overwhelming evidence for a model with time effects. Most importantly, by testing for AR(1), we cannot find residual autocorrelation in our first-differenced data, suggesting that our dynamic specification does not suffer from endogeneity. Applying the panel-corrected standard errors in this situation, as elaborated in Table 4, seems acceptable. Differencing the data wipes out the country effects. Thus, the endogeneity bias caused by the correlation between the lagged dependent variable and the country effects vanishes (see Hsiao 2003: 85–86). In order to take into account the remaining endogeneity due to correlation of the lagged dependent variable and the error term, it is usually suggested applying more consistent estimation procedures (such as IV or GMM; see Baltagi 2001: 131; Wawro 2002). However, Monte Carlo simulations have shown that, for an unbalanced panel, as in our case, *and* T = 30, the FE performs just as well as

Table 4. Dynamic specification in first differences: Fixed effects, correction for heteroskedasticity and cross-sectional correlation (dependent variable: total government expenditure as percentage of GDP, first differences)

	FE(CT)	FE(T)	WLS	PCSE
Δ Government expenditure$_{t-1}$	0.188*** (0.048)	0.226*** (0.047)	0.266*** (0.044)	0.226*** (0.068)
Δ Unemployment	0.423*** (0.080)	0.401*** (0.079)	0.438*** (0.069)	0.401*** (0.091)
Δ GDP/capita growth	–0.206*** (0.026)	–0.213*** (0.026)	–0.216*** (0.023)	–0.213*** (0.033)
Δ Dependency ratio	0.452** (0.222)	0.436*** (0.206)	0.348** (0.157)	0.436** (0.213)
Left cabinet portfolios	–0.249 (0.197)	0.040 (0.145)	0.154 (0.118)	0.039 (0.164)
Christian Democrat portfolios	–0.629 (0.459)	–0.069 (0.208)	–0.042 (0.161)	–0.069 (0.212)
Δ Trade	–0.023 (0.021)	–0.027 (0.021)	–0.031* (0.018)	–0.027 (0.026)
Δ Low wage imports	0.014 (0.035)	0.020 (0.035)	0.025 (0.028)	0.020 (0.038)
Δ Foreign direct investment	–0.134* (0.072)	–0.120* (0.071)	–0.130** (0.063)	–0.120 (0.082)
R^2	0.520	0.461	0.589	0.428
N (observations)	512	512	512	512
F (Country effects)	1.01	–	–	–
F (Time effects)	2.92***	2.86**	–	–
F (Country and time effects)	2.22***	–	–	–
Cross-sectional correlation	162.22*	161.74*	–	–
Groupewise heteroskedasticity	362.37***	440.48***	–	–
Autocorrelation (AR1)	1.30	1.32	4.49**	1.32

Notes: see Table 1; constant and fixed effects not reported; WLS = GLS-transformed estimator taking into account groupwise heteroskedasticity (Greene 2000: 594–599); White = OLS model with White's (1980) heteroskedasticity-robust standard errors; PCSE = OLS with Beck and Katz's (1995) panel-corrected standard errors.

or better than the alternatives (see Judson & Owen 1999: 13). This is an additional justification for our specification.

In all specifications, the coefficient of the lagged dependent variable is stable and around 0.25, suggesting a persistency of public-spending decisions. In most of the industrialized countries, current spending is widely determined by indisposable positions and legal admissions (e.g., wages and salaries of public employees), which, in turn, makes this parameter estimate plausible.

Robustness and stability analysis

By estimating a single coefficient for the whole period, Garrett and Mitchell assume a constant and stable globalization-welfare state nexus. However, the impact of both globalization and political pressures on government spending may well change over time. In addition, in applied work, one is often confronted with structural breaks, which may be caused by economic shocks (e.g., the oil crisis in the 1970s) or simply by changing statistical conventions. To account for these considerations, we re-estimate the first differenced, fixed time effects model of Table 4 with interaction dummies for three subperiods: 1964–1973 (up to the first oil crisis), 1974–1983 (characterized by growing unemployment and inflation, as well as decreasing growth rates), and finally, 1984–1993, a period of stable economic growth and, according to conventional wisdom, the emergence of globalization. We interact these period dummies with our variables of interest (i.e., both the partisan composition of government and the globalization variables). The results are presented in Table 5.

Compared to our favoured specification in Table 4 (FE(T)/PCSE), the coefficients of the core economic variables, unemployment, economic growth and the dependency ratio, are almost unchanged. Most important, however, is that by applying the likelihood ratio (LR) tests on the (unrestricted) specification with period-specific coefficients for the core variables, we get results indicating a rejection of the null of significant differences between the period-specific effects. In other words, the association between government expenditure and partisan composition as well as globalization remains almost stable over time.[12] This finding is confirmed by the t statistics of the coefficients: Again, our core economic variables are significant, whereas the only significant effects concerning partisanship and globalization are those of trade and FDI, both in the period 1984 to 1993. It is worth noting that the coefficient on FDI reversed its sign between the late 1970s and the 1980s. From a statistical point of view, this may be due to structural breaks inherent in these series. Substantively, this may be indicative of an impact of economic

Table 5. Dynamic specification in first differences: Period-specific slopes (dependent variable: total government expenditure as percentage of GDP, first differences)

	Estimate	s.e.	LR statistic
Δ Government expenditure$_{t-1}$	0.230***	0.047	
Δ Unemployment	0.412***	0.080	
Δ GDP/capita growth	–0.207***	0.026	
Δ Dependency ratio	0.458**	0.223	
Left cabinet portfolios, 1963–1973	0.146	0.250	Left cabinet
Left cabinet portfolios, 1974–1983	–0.178	0.249	portfolios:
Left cabinet portfolios, 1984–1993	0.133	0.264	$\chi 2(2) = 1.15$
Christian Democrat portfolios, 1963–1973	–0.178	0.354	Christian Democrat portfolios:
Christian Democrat portfolios, 1974–1983	0.164	0.389	$\chi 2(2) = 0.81$
Christian Democrat portfolios, 1984–1993	–0.277	0.364	
Δ Trade, 1963–1973	0.001	0.053	Δ Trade:
Δ Trade, 1974–1983	–0.010	0.031	$\chi 2(2) = 1.82$
Δ Trade, 1984–1993	–0.060*	0.033	
Δ Low wage imports, 1963–1973	0.080	0.061	Δ Low wage
Δ Low wage imports, 1974–1983	–0.029	0.060	imports:
Δ Low wage imports, 1984–1993	0.015	0.061	$\chi 2(2) = 1.82$
Δ Foreign direct investment, 1963–1973	–0.062	0.217	Δ Foreign direct
Δ Foreign direct investment, 1974–1983	0.139	0.211	investments:
Δ Foreign direct investment, 1984–1993	–0.157*	0.083	$\chi 2(2) = 1.94$
R^2	0.512		
N (observations)	512		
F (Time effects)	2.47***		
LR (period-specific slopes), $\chi 2(10)$	9.09		
Cross-sectional correlation	153.77		
Groupewise heteroskedasticity	387.60***		
Autocorrelation (AR1)	1.52		

Notes: see Table 1; constant and fixed effects not reported.

internationalization: on average, an increase in FDI is associated with an increase in government spending during the late 1970s and early 1980s, which suggests that the compensation hypothesis was salient at that time. By contrast, the efficiency hypothesis tends to be confirmed for the period since the mid-1980s. Hence the efficiency versus compensation debate might be resolved by the proposition that governments first attempted to compensate their citizens for greater exposure to the world market, but later had to succumb to the pressures from growing deficits and debt as well as from economic internationalization. However, this is an ex-post and ad-hoc interpretation based on rather weak findings that we do not wish to over-emphasize.

Finally, as emphasized above, our attention should be devoted to the robustness of our results against the impact of cross-sectional outliers. To discover such effects, we perform a Jacknife analysis (Efron & Tibshirani 1993: Chapter 11) on our favoured FE(T)/PCSE specification of Table 4. More precisely, we re-estimate the model by excluding one country after the other. The resulting minimum and maximum values of the point estimates are presented in Table 6. The minimum values as well as the excluded country for which that estimate was obtained are shown in the first double column, and the maximum ones in the last double column. The point estimates of the entire sample (taken from Table 4) are reported in the middle column. A glance at the partisanship variables shows us that the coefficient estimates strongly depend on the countries included. The most extreme lower and upper deviations from the estimated coefficient are caused by the exclusion of France and Sweden for leftist governments and by the exclusion of Belgium and Germany for Christian Democratic governments. The other coefficients are much more stable. None of them, except trade, changes sign due to the exclusion of a particular country. The most extreme coefficient estimates of the domestic socioeconomic factors

Table 6. Jackknife analysis (dependent variable: total government expenditure as percentage of GDP, first differences)

	Minimum	Country	Estimate	Maximum	Country
Δ Government expenditure$_{t-1}$	0.213	DK	0.226	0.245	IT
Δ Unemployment	0.368	CA	0.401	0.444	IE
Δ GDP/capita growth	–0.234	IE	–0.213	–0.199	IT
Δ Dependency ratio	0.397	SE	0.436	0.501	NL
Left cabinet portfolios	–0.005	FR	0.040	0.124	SE
Christian Dem. portfolios	–0.143	BE	–0.069	0.024	DE
Δ Trade	–0.040	IT	–0.027	0.003	IE
Δ Low wage imports	0.007	NO	0.020	0.046	AU
Δ FDI	–0.182	NO	–0.120	–0.065	SE

Notes: Entries are coefficient estimates from FE(T) in Table 4, together with minimum and maximum coefficient estimates resulting from re-estimates of the model while excluding each country one at a time. They reveal the responsiveness of the coefficient estimates to the inclusion of particular countries.

are within approximately a 10 per cent range of the coefficient estimate calculated by using the full set of countries. The economic variables relating to the external effects (trade, imports from low-wage countries and FDI) are less stable – deviations of 50 to 100 per cent of the coefficient estimate are caused by the exclusion of a single country – which gives additional salience to the inference derived from the lack of significance that there is no clear effect. These findings indicate that decisions on government expenditure tend to be driven by the domestic socioeconomic environment to a much greater extent than by political factors or issues relating to globalization (as claimed by Garrett and Mitchell 2001).

Summary and conclusions

It has become common practice in the empirical analysis of political economy to use panel data for drawing inferences on a wide range of research topics. In fact, data that combine cross-sectional and time-series information have given us new insights into the course of political processes. However, despite these merits, our analysis calls for great caution in using panel data in comparative political economy. Exemplified by the controversy over the impact of globalization and the partisan composition of government on total government expenditure, we identify a lack of concern about consistent empirical specifications. In particular, we have re-estimated a model by Garrett and Mitchell (2001).

Our findings may be summarized as follows. First, in general, panel data inferences are sensitive to the model specification. In our context, the underlying data are highly persistent, and therefore the specification has to focus on the time pattern of the data. Second, introducing a lagged dependent variable to account for autocorrelation causes additional problems, in particular if autocorrelation remains after correcting for it. Most importantly, the estimates become biased due to endogeneity. Our proposition for this particular model is to specify an autoregressive model in first differences with time effects, the latter taking into account common shocks. By testing for remaining violations of the panel assumptions, we end up with a statistically acceptable model that allows us to apply panel-corrected standard errors as proposed by Beck and Katz (1995), although these hardly differ from OLS standard errors – despite the fact that the tests for country-specific heteroskedasticity and cross-sectional correlation suggested the presence of such a residual structure. Apparently, these problems matter less in this particular model.

Third, it is rather unlikely that social phenomena, such as the globalization-welfare state nexus considered by Garrett and Mitchell (2001) and subsequent studies, are stable over time. We take into account this argument by specifying a model with period-specific effects of the core variables, but we find no convincing evidence that the lack of statistically significant estimates in the model for the full period analyzed is due to systematic changes in the coefficients over time. However, the restricted number of observations from which each of these parameters is estimated does not allow us to make a definite judgement.

In contrast to Garrett and Mitchell (2001), our empirical analysis indicates that neither globalization nor the partisan composition of government play a decisive role in explaining cross-country variation in the dynamics of government expenditure. Hence, we find a confirmation neither of the efficiency hypothesis nor of the compensation hypothesis, although we find weak evidence indicating that initial attempts at compensating for international risks seem to have been replaced by an ubiquitous move to efficiency. Instead, most parts of the dynamics in government expenditure are explained by the domestic economic environment such as growth, unemployment or the dependency ratio (for a similar conclusion, see Castles 2001). Some of these may be affected by partisan politics or economic globalization, but this is a different story. Therefore, while we have focused on the statistical model used by Garrett and Mitchell (2001), much could be said about the lack of theoretical foundations in their empirical model and the way in which globalization interacts with politics to cause particular outcomes, among which is the development of government expenditure dynamics.

Substantively, the shift from a model in levels to a model in first differences implies that the object of analysis has changed from long-term differences (i.e., the level of public expenditure) to short-term adjustments (i.e., the growth of public expenditure). We should not be too surprised that an analysis of short-run adjustments does not support propositions about long-run effects. Furthermore, if we compare the results of a purely cross-sectional 'between model' in levels with that of our first differenced specification, we can draw a second major conclusion. While the former clearly indicates a positive effect of leftist cabinet portfolios and a somewhat more ambiguous positive effect of Christian Democratic portfolios on the levels of public expenditure, the latter rejects any partisan effects on the first differences. However, this result is not necessarily a contradiction. Despite the shifting nature of cabinet portfolios, countries with an historically strong social democracy (e.g., Sweden, Denmark, Austria) or Christian democracy (e.g., Germany) or, conversely, a strong liberal tradition (e.g., United States, Canada), tend to have prolonged periods with high values on these variables during the period analyzed, too. Therefore, the country means of government composition in the between-model practically serve as a proxy for the historical power constellation. Since all countries exhibited an upward trend in public expenditure during the period analyzed, the differences in mean public expenditure are accounted for by the government composition variables. However, no systematic variation in growth is observable just because all countries experienced the upward trend.

Finally, what do our findings tell us about the usefulness of panel analysis for the type of questions asked by comparative political economists? We would like to give our answer in three steps. First, the either nonstationary or non-varying character of many core variables in comparative political economy constitute a serious impediment to a sensible interpretation of coefficient estimates in pooled models. In macroeconomics, tools like cointegration and error correction models have been developed to disentangle short- and long-run effects in the context of nonstationary single time series. However, since these approaches focus on the time dimension of a data set, they are less suitable for analyzing the impact of institutions across countries, which is at the centre of interest in comparative political economy. We contend that those who are interested in the consequences of cross-country differences in political institutions – that by definition exhibit little short-term variation – should take the fact that panel data do have a longitudinal dimension much more into consideration. Neither persistency nor dynamics should be regarded as nuisances to be corrected. Quite the contrary: the longitudinal dimension of panel data makes this type of data ill-suited for answering research questions posed in a strictly cross-sectional context. The solution is not to attach a variety of bells and whistles to the model in order to correct for autocorrelation and other 'panel nuisances', but to pay much more attention to the short-term dynamics themselves.

Second, long-term variation in levels is always the result of systematic variation in short-term dynamics. This points to an interesting avenue for future research: the theoretical debate about partisan effects on short-term adjustments in government expenditure is notoriously underdeveloped and sometimes even dismissed as ill-guided (Huber & Stephens 2001: 36–38). However, if persistence in time series has any substantive meaning in the context of government expenditure, it is that political actors usually change expenditure programmes only at the margin. If one believes, for instance, that the partisan composition affects the long-run political outcome, it is caused by marginal changes in programme designs, the effect of which accumulates over time. Therefore it is worthwhile to study not only the long-run developments, but also the short-run political-economic processes. An important contribution in this regard would be the development of theories of short-term adjustments on which empirical analysis could be based. Such an approach would need to derive much more specific hypotheses from underlying theories of strategic interaction that have been developed in recent years in the fields of political economy and public choice (e.g., Persson & Tabellini 2000). It prohibits justifying the inclusion of variables in statistical models with simplistic claims in the style of 'leftist parties are associated with higher public spending because they want to enlarge the welfare state'. In contrast, it would be necessary to specify explicitly the framework conditions under which the association holds and to make a judgement on the expected effect size and timing (including the time lag and adjustment process). In terms of statistical modeling, this approach opens the avenue to focusing on associations in first differences.

Third, it is an almost forgotten truism of comparative analysis that if a research question definitely refers to cross-sectional differences in levels, it makes sense to answer that question by emphasizing the cross-sectional dimension despite the seductive but treacherous glitter of significance stars in pooled regressions. This means that purely cross-sectional regression analyses still have their own value. Since they are not disturbed by the nuisances inherent in time series they can be used as validating instruments for pooled analyses. Particularly in situations in which the focused variables vary too little over time to be relevant at the conceptual level, there is not much to be gained by pooling repeated observations over time (see also Jackman 1985: 173–175; King et al. 1994: 48). Instead, the increase in leverage obtained by the multiplication of the number of observations can be used for exploring the stability of an empirical association over time.

In sum, our exploration has led us to take on the position that pooling is no panacea. There are specific situations in which pooling the data indeed makes much sense – for example, if there are reasons for testing an effect assumed to be equal across space and time or in situations where the short-term adjustments are the main focus of interest. Pooled analysis is less suited if the research question refers to long-term effects of level differences because, in that context, the statistical problems inherent in the time dimension overwhelm the analysis. However, in more situations than might be acknowledged at the moment it may prove justified to focus on short-term adjustments that might more realistically capture the relevant time frame of political decision-making processes, despite the fact that many outcomes materialize only in the long run.

Acknowledgments

Major parts of this article were written while Bernhard Kittel was a research fellow at the Max Planck Institute for the Study of Societies in Cologne during a visit of Hannes Winner in November 2001. We wish to express our thanks to the Institute for its welcome and support. Furthermore, we wish to thank Andreas Broscheid, Francis Castles, Peter Egger, Markus Freitag, Steffen Ganghof, Philipp Rehm and Fritz Scharpf for their useful comments. We are also grateful to Hans Keman and two anonymous referees for their suggestions.

Notes

1. A practical reason is that the data set made available on the Internet by Geoffrey Garrett did not contain the data for covered interest rate differentials, which are used as additional proxy for globalization. Garrett and Mitchell (2001) refer to data compiled by Shepherd (1994). Unfortunately, this data set contains fewer observations than they have used for estimating their model for social expenditure, at least as far as can be inferred from the number of observations mentioned in Table 5 (see Garrett & Mitchell 2001: 166).
2. The rejection of poolability means to sacrifice the feature that most comparative political economists regard as one of the prime advantages of panel data: the (often only apparent) increase in degrees of freedom obtained by repeating cross-sectional observations over time (see Hsiao 1986: 2). Since much effort has been put into the collection of panel data sets, and practically all published contributions to comparative political economy using panel data assume poolability by fiat, we simply note that this might be problematic.
3. A Hausman test (Hausman 1978) additionally justifies the fixed effects approach instead of a random effects model, which assumes uncorrelatedness of the country effects and the regressors (see Baltagi 2001: 15).
4. Similarly, Maddala (1999: 434) concludes: 'Thus, choosing the countries used in the panel study is as crucial as (if not more than) the choice of the estimation method.'
5. Garrett and Mitchell (2001: 163) emphasize that country effects should be included in order to capture idiosyncrasies and that any time-constant variable should be regarded as being part of the 'underlying historic fabric of a country'. This view is not convincing since one of the main interests of political economists in this kind of quantitative analysis is to determine whether institutional variables capture cross-sectional variation to an extent that makes the inclusion of country dummies unnecessary (see Beck & Katz 2001, in the context of limited dependent variables).
6. We were unable to reproduce their results exactly, although we used the data set published on Garrett's webpage. The most serious deviation is the change in sign of the FDI variable. One possible reason is that the procedure they used (xtpcse in STATA 6) may be slightly different from ours (xtpcse in STATA 7). However, we believe that our estimates are close enough to proceed with our analysis. Note that we shifted the partisanship effects by a factor 100 in order to obtain nicer table entries.
7. Usually, this type of bias is accounted for by an instrumental variables (IV) or a methods of moments (GMM) approach (see Baltagi 2001: 131–153; Wawro 2002). However, since we are additionally confronted with serial correlation, these solutions would no longer work.
8. Technically, an autoregressive coefficient above unity violates the *invertability condition* that is needed to reassure stable solutions in a stochastic process (see Maddala & Kim

1998: 13). Consequently, the conventional unit roots tests turn out to be invalid. See Elliott and Stock (2001) for tests on autoregressive coefficients near one.

9. In principle, all variables in the model have to be stationary. For reasons of space, we do not present and discuss the test results for the independent variables since we will decide below to first difference the data anyway.
10. A further possibility is to estimate an error correction model (ECM) which includes either the stationary residuals from a preliminary levels regression or the levels of the cointegrating variables. In contrast to the practice in comparative political economy (e.g., Iversen & Cusack 2000), however, we insist on the need to introduce the levels variables only on the basis of a test for cointegration because we otherwise might regress a stationary variable (the first differenced dependent variable) on a set of nonstationary variables (the independent variables in levels). In this regard, the ECM is confronted with the same practical problems as a fully developed cointegration model.
11. A more technical and practical reason is that, in general, time series with $T \cong 30$ are too short for the estimation of reliable parameters in the cointegration framework (see Maddala & Wu 1999).
12. A glance at Table 6 reveals that the coefficients change signs across the periods. However, the size is far too small in relation to the standard errors to warrant any conclusion to the contrary.

References

Achen, C.F. (2000). Why Lagged Dependent Variables can Suppress the Explanatory Power of Other Independent Variables. Paper presented at the Annual Meeting of the Political Methodology Section of the American Political Science Association, Los Angeles, 20–22 July.

Baltagi, B.H. (2001). *Econometric analysis of panel data*, 2nd edn. New York: Wiley.

Beck, N. & Katz, J. (1995). What to do (and not to do) with time-series cross-section data. *American Political Science Review* 89(3): 634–647.

Beck, N. & Katz, J. (1996). Nuisance vs. substance: Specifying and estimating time-series cross-section models. *Political Analysis* 6: 1–34.

Beck, N. & Katz, J. (2001). Throwing out the baby with the bath water: A comment on Green, Kim and Yoon. *International Organization* 55(2): 487–496.

Bretschger, L. & Hettich, F. (2002). Globalisation, capital mobility and tax competition: Theory and evidence for OECD countries. *European Journal of Political Economy* 18(4): 695–716.

Cameron, D. (1978). The expansion of the public economy: A comparative analysis. *American Political Science Review* 72(4): 1243–1261.

Castles, F.C. (2001). On the political economy of recent public sector development. *Journal of European Social Policy* 11(3): 195–211.

Daveri, F. & Tabellini, G. (2000). Unemployment, growth and taxation in industrial countries. *Economic Policy* 15(30): 49–104.

Efron, B. & Tibshirani, R.J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Elliott, G. & Stock, J.H. (2001). Confidence intervals for autoregressive coefficients near one. *Journal Of Econometrics* 103(1/2): 155–181.

Garrett, G. (1995). Capital mobility, trade and the domestic politics of economic policy. *International Organization* 49(4): 657–687.
Garrett, G. (1998). *Partisan politics in the global economy*. Cambridge: Cambridge University Press.
Garrett, G. & Mitchell, D. (2001). Globalization, government spending and taxation in the OECD. *European Journal of Political Research* 39(1): 145–177.
Greene, W. (2000). *Econometric analysis* 4th edn. Englewood Cliffs, NJ: Prentice Hall.
Hausman, J.A. (1978). Specification tests in econometrics. *Econometrica* 46(6): 1251–1271.
Hausman, J.A. & Taylor, W.E. (1981). Panel data and unobservable individual effects. *Econometrica* 49(6): 1377–1398.
Hicks, A.M. & Swank, D. (1992). Politics, institutions and welfare spending in industrialized democracies, 1960–1982. *American Political Science Review* 86(3): 658–674.
Hsiao, C. (2003). *Analysis of panel data*, 2nd edn. Cambridge: Cambridge University Press.
Huber, E. & Stephens, J. (2001). *Development and crisis of the welfare state*. Chicago, IL: University of Chicago Press.
Im, K.S., Pesaran, H. & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115(1): 53–74
Iversen, T. & Cusack, T. (2000). The causes of welfare state expansion: De-industrialization or globalization? *World Politics* 52(3): 313–349.
Jackman, R. (1985). Cross-national statistical research and the study of comparative politics. *American Journal of Political Science* 29(1): 161–182.
Judson, R.A. & Owen, A.L. (1999). Estimating dynamic panel data models: A guide for macroeconomists. *Economics Letters* 65(1): 9–15.
Kao, C. & Chiang, M.-H. (2000). On the estimation and inference of a cointegrated regression in panel data. *Advances in Econometrics* 15: 179–222.
Katzenstein, P.J. (1985). *Small states in world markets: Industrial policy in Europe*. Ithaca, NY: Cornell University Press.
King, G., Keohane, R.O. & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.
Kittel, B. & Winner, H. (2002). How Reliable is Pooled Analysis in Political Economy? The Globalization-Welfare State Nexus Revisited (Discussion Paper 03/02). Cologne: Max Planck Institute for the Study of Societies.
Kiviet, J.F. (1995). On bias, inconsistency and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68(1): 53–78.
Kmenta, J. (1990). *Elements of econometrics*. London: Macmillan.
Levin, A. & Lin, C.-F. (1992). Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties (Discussion Paper 92–23). San Diego, CA: University of California.
Maddala, G.S. (1999). On the use of panel data methods with cross-country data. *Annales d'Economie et de Statistique* 55/56: 429–449.
Maddala, G. S. & Kim, I.-M. (1998). *Unit roots, cointegration and structural change*. Cambridge: Cambridge University Press.
Maddala, G.S. & Wu, S. (1999). A comparative study of unit roots tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics* 61(4): 631–652.
Moses, J.W. (2001). A Methodological Critique of Statistical Studies of Globalization. Unpublished manuscript, Norwegian University of Science and Technology (NTNU), Trondheim.
Mueller, D.C. (2003). *Public choice III*. Cambridge: Cambridge University Press.
Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica* 49: 1417–1426.

Pedroni, P. (2000). Fully modified OLS for heterogeneous cointegrated panels with multiple regressors. *Advances in Econometrics* 15: 93–130.

Persson, T. & Tabellini, G.E. (2000). *Political economics: Explaining economic policy*. Cambridge, MA: MIT Press.

Pierson, P. (1996). The new politics of the welfare state. *World Politics* 48(2): 143–179.

Quinn, D. (1997). The correlates of change in international financial regulation. *American Political Science Review* 91(3): 531–551.

Rodrik, D. (1997). *Has globalization gone too far?* Washington, DC: Institute for International Economics.

Rodrik, D. (1998). Why do more open economies have bigger governments? *Journal of Political Economy* 106(5): 997–1032.

Schulze, G.G. & Ursprung, H.W. (1999). Globalisation of the economy and the nation-state. *The World Economy* 22(3): 295–352.

Shepherd, W.F. (1994). *International financial integration: History, theory and applications in OECD countries*. Aldershot: Avebury.

Swank, D. (1998). Funding the welfare state: Globalization and the taxation of business in advanced market economies. *Political Studies* 46(4): 671–692.

Swank, D. & Steinmo, S. (2002). The new political economy of taxation in advanced capitalist democracies. *American Journal of Political Science* 46(3): 642–655.

Wawro, G. (2002). Estimating dynamic panel data models in political science. *Political Analysis* 10(1): 25–48.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48(4): 817–838.

Correspondence to: Bernhard Kittel, University of Amsterdam, Amsterdam School for Social Science Research, Oudezijds Achterburgwal 185, 1012 DK Amsterdam, The Netherlands.
Tel.: ++31 20 525 7375; Fax: ++31 20 525 3010; E-mail: B.E.A.Kittel@uva.nl