A SPECIAL COVARIANCE STRUCTURE FOR RANDOM COEFFICIENT MODELS WITH BOTH BETWEEN AND WITHIN COVARIATES

Kurt S. Riedel

IPP 5/36 July 1990

MAX-PLANCK-INSTITUT FÜR PLASMAPHYSIK

8046 GARCHING BEI MÜNCHEN

MAX-PLANCK-INSTITUT FÜR PLASMAPHYSIK

GARCHING BEI MÜNCHEN

A SPECIAL COVARIANCE STRUCTURE FOR RANDOM COEFFICIENT MODELS WITH BOTH BETWEEN AND WITHIN COVARIATES

Kurt S. Riedel

IPP 5/36 July 1990

Guest Scientist from Courant Institute of Mathematical Sciences
New York University, New York, N.Z. 10012, USA

Die nachstehende Arbeit wurde im Rahmen des Vertrages zwischen dem Max-Planck-Institut für Plasmaphysik und der Europäischen Atomgemeinschaft über die Zusammenarbeit auf dem Gebiete der Plasmaphysik durchgeführt.

A SPECIAL COVARIANCE STRUCTURE FOR RANDOM COEFFICIENT MODELS WITH BOTH BETWEEN AND WITHIN COVARIATES

K.S. Riedel
Courant Institute of Mathematical Sciences
New York University
New York, N.Y. 10012

Abstract

We review random coefficient (RC) models in linear regression and propose a bias correction to the maximum likelihood (ML) estimator. Asymptotic expansions of the ML equations are given when the between individual variance is much larger or smaller than the variance from within individual fluctuations. The standard model assumes all but one covariate varies within each individual, (we denote the within covariates by $\vec{x}_1$). We consider random coefficient models where some of the covariates do not vary in any single individual (we denote the between covariates by $\vec{x}_o$). The regression coefficients, $\vec{\beta}_k$, can only be estimated in the subspace $\mathbf{X}_k$ of $\mathbf{X}$. Thus the number of individuals necessary to estimate $\vec{\beta}$ and the covariance matrix $\underline{\underline{\Delta}}$ of $\vec{\beta}$ increases significantly in the presence of more than one between covariate. When the number of individuals is sufficient to estimate $\vec{\beta}$ but not the entire matrix $\underline{\underline{\Delta}}$, additional assumptions must be imposed on the structure of $\underline{\underline{\Delta}}$. A simple reduced model is that the between component of $\vec{\beta}$ is fixed and only the within component varies randomly. This model fails because it is not invariant under linear

coordinate transformations and it can significantly overestimate the variance of new observations. We propose a covariance structure for $\underline{\underline{\Delta}}$ without these difficulties by first projecting the within covariates onto the space perpendicular to the between covariates.

KEYWORDS:Random Coefficient Models, Special Covariance Structures, Mixed Models, Estimated Generalized Least Squares

I. Introduction

We consider structured linear regressions of the form, $y = \vec{\beta} \cdot \vec{x}$, where the data is divided into discrete subgroups which we term individuals. When the data from a single individual contains variation in all covariate directions, an estimate for the regression coefficient vector can be constructed purely from that individual. We denote the estimate arising from the kth individual by $\hat{\vec{\beta}}_k$.

In many physical, biological, and economic systems, the regression coefficients, $\vec{\beta}_k$, vary from individual to individual. If the differences in the parameter vectors, $\vec{\beta}_k$, are due to many small differences in the individuals, instead of one or two significant latent variables, a statistical treatment is still possible. In these cases, the random coefficient (RC) model[1-9], introduced by Swamy, is applied. Swamy assumes that the parameter vectors, $\vec{\beta}_k$, are randomly distributed about a mean vector $\vec{\beta}$. More precisely, $\vec{\beta}_k = \vec{\beta} + \delta\vec{\beta}_k$, where $E[\delta\vec{\beta}_k] = 0$, $E[\delta\vec{\beta}_k \delta\vec{\beta}_k^t] = \underline{\underline{\Delta}}$, and $E[\delta\vec{\beta}_k \delta\vec{\beta}_l^t] = 0$, $k \neq l$. The random coefficient matrix, $\underline{\underline{\Delta}}$, is usually estimated from the residuals, $\hat{\vec{\beta}}_k - \hat{\vec{\beta}}$.

The random coefficient matrix, $\underline{\underline{\Delta}}$, determines the error covariance matrix of the individual observations, $\underline{\underline{\Sigma}}$, as expressed by Eq. (3). The random coefficient matrix, $\underline{\underline{\Delta}}$, and the resulting covariance matrix $\underline{\underline{\Sigma}}$, serve three main purposes. First, more efficient estimation of $\vec{\beta}$ is obtained using GLS regression with $\underline{\underline{\Sigma}}$ specified by Eq. (4,6). Second, to test hypotheses for additional variables, one must use $\underline{\underline{\Sigma}}^{-1}$ as the metric. Third, the RC matrix enables one to estimate the variance of the mean of the $\overline{y}_k$ observations, In

the limit as the number of observations for a the kth individual tends to infinity, this single individual variance reduces to $\vec{\vec{x}}^t \underline{\underline{\Delta}} \vec{\vec{x}}$. In general, hypothesis testing requires the actual covariance structure of the errors to be modeled to extremely high accuracy to be useful.

The original work of Swamy assumed that all the regression coefficients can be determined in each individual. *We consider situations where not all the covariates can be varied in the individual experiments.* One such problem is the extrapolation of performance parameters in controlled nuclear fusion experiments[6–8]. A database consisting of a large number of datapoints from eight fusion experiments has been collected. The energy confinement time, a performance parameter, is assumed to be a log linear function of certain bulk variables such as plasma current, magnetic field and device size. However, several of these independent variables, namely the device size and device shape, cannot be varied in the individual experiments. We allow the variation of the covariates to be strongly correlated.

We call the covariates which are fixed for a given individual "between variables", denoted by $\vec{x}_0$. The covariates which vary in each individual are called "within variables", denoted by $\vec{x}_1$. We consider the case where the within regression parameters, $\vec{\beta}_1$, can be reasonably accurately determined for each individual. We assume that the variation in the within regression coefficients can be treated with the Swamy random coefficient model.

We note that only $\vec{\beta}_o \cdot \vec{\vec{x}}_o$ and not the entire $\vec{\beta}_o$ can be estimated for a single individual. Thus substantially more individuals are necessary to estimate $\vec{\beta}$ and $\underline{\underline{\Delta}}$. Since $\vec{x}_0$ is not varied in a given individual, $\vec{\beta}_0$ can only

be determined using an estimated general least squares (EGLS) estimate on all individuals simultaneously. We concentrate on cases where the number of individuals is only slightly larger than m_o, the number of regression parameters which can only be determined by interindividual comparisons. *In this article, we propose a special covariance structure for the random coefficient matrix* $\underline{\underline{\Delta}}$ *when only a submatrix, corresponding to the within covariates,* $\underline{\underline{\Delta}}_1$*, can accurately be determined using the standard RC variance estimators.*

In predicting the variance of $\overline{y}_k$, we wish to choose the model of $\underline{\underline{\Delta}}$ which minimizes the heteroscedasticity of $\underline{\underline{\Delta}}$ among models which fit the data equally well. The within covariates can be reparametrised by an arbitrary linear transformation of the form $\vec{x}_1' = \vec{x}_1 + \underline{\underline{A}}\,\vec{x}_0$. To minimise the known heteroscedasticity, we choose $\underline{\underline{\Delta}}$ to enforce

$$\vec{x}^t \underline{\underline{\Delta}} \vec{x} = (\vec{x}_1 - \underline{\underline{A}}\,\vec{x}_o)^t \underline{\underline{\Delta}}_1 (\vec{x}_1 - \underline{\underline{A}}\,\vec{x}_o) + \sigma_o^2. \tag{1}$$

where $\underline{\underline{A}}$ minimizes $\left|\vec{x}_1 + \underline{\underline{A}}\,\vec{x}_0\right|^2$ averaged over the sample.

This covariance structure has the following desirable properties. First, the variance structure is invariant under linear changes of variable. Second, the predicted variance in the mean values of the distinct individuals is more nearly homoscedastic, i.e. independent of the choice of individual. This reduction in heteroscedasticity applies not only to the existing sample, but also to future observations, provided that the new samples have the same correlations in the $\vec{x}$ covariates as the existing data.

In section 2, we review the Swamy random coefficient model for the stan-

dard case that all covariates are varied in all individuals. Two new results are presented, a bias correction for the maximum likelihood (ML) estimator and asymptotic expansions of the ML equations when $\underline{\underline{\Delta}} >> \sigma_k^2 \underline{\underline{X}}_k^t \underline{\underline{X}}_k^{-1}$ or $\underline{\underline{\Delta}} << \sigma_k^2 \underline{\underline{X}}_k^t \underline{\underline{X}}_k^{-1}$. In section 3, we discuss the problems associated with between covariates in the random coefficient model and a simple but flawed solution is given. In section 4, we propose a particular RC structure, which is defined by minimizing the average predicted heteroscedasticic variance, (the first term of Eq. 1), with respect to the matrix $\underline{\underline{A}}$. We then discuss the advantages of this covariance structure and of selecting the most homoscedastic error structure. In appendix A, the invariance of our random coefficient model under linear transformations is demonstrated.

The proposed error structure is degenerate with a nullspace of rank $m_o - 1$. When $\underline{\underline{\Delta}} >> \sigma_k^2 \underline{\underline{X}}_k^t \underline{\underline{X}}_k^{-1}$, this leads to the inversion of illconditioned matrices. Thus a biased estimation technique such as ridge regression may be necessary. In appendix B, a number of standard results of ridge regression are presented for the case of an arbitrary covariance matrix $\underline{\underline{\Sigma}}$ and perturbing ridge matrix $\underline{\underline{P}}$.

II. Random Coefficient Models: All Covariates varied in Each Individual

We begin our review of Swamy's random coefficient model[1-9] by establishing notation. Our dataset consists of N distinct individuals and the kth individual has n_k independent samples. We use the letters, k and l to index individuals and the letters, i and j to index samples of a given

individual. Thus the entire dataset can be identified with a double index, $\{k, i\}$.

We assume the kth individual is a linear function of m covariates, $\vec{x}$. The errors in the linear model for the kth individual are assumed to be independent with a constant variance σ_k^2. Thus the linear model for the kth individual is

$$\vec{y}_k = \underline{\underline{X}}_k \vec{\beta}_k + \vec{e}_k$$

where $\vec{y}_k$ and $\vec{e}_k$ are $(n_k \times 1)$ vectors and $\underline{\underline{X}}_k$ is the $(n_k \times m)$ data matrix $\vec{e}_k$ is distributed $E[\vec{e}_k, \vec{e}_l{}^t] = \sigma_k^2 \delta_{kl} I_{n_k}$. We denote the mean values of the individuals as $\overline{y}_k$ and $\vec{\overline{x}}_k$. In practice, all covariates are usually centered about $\overline{\overline{y}} = 1/N \sum \overline{y}_k$, and $\vec{\overline{x}} = 1/N \sum \vec{\overline{x}}_k$.

In the Swamy model, we assume that the regression coefficients for each individual, $\vec{\beta}_k$ have an *apriori* random distribution with $E[\vec{\beta}_k] = \vec{\beta}$ and

$$E[(\vec{\beta}_k - \vec{\overline{\beta}})(\vec{\beta}_l - \vec{\overline{\beta}})^t] = \underline{\underline{\Delta}} \delta_{kl} \tag{2}$$

where $\underline{\underline{\Delta}}$ is an $m \times m$ covariance matrix.

This random coefficient model implies that the covariance matrix for the observation errors has a block structure. Furthermore, the error matrix for the k-th block reduces to

$$\underline{\underline{\Sigma}}_k = \sigma_k^2 \underline{\underline{I}}_k + \underline{\underline{X}}_k \underline{\underline{\Delta}} \underline{\underline{X}}_k^t \tag{3}.$$

Given the covariance, $\underline{\underline{\Delta}}$, the parameter vector $\vec{\beta}$ may be estimated using the GLS estimator:

$$\hat{\vec{\beta}} = (\underline{\underline{X}}^t \underline{\underline{\Sigma}}^{-1} \underline{\underline{X}})^{-1} \underline{\underline{X}}^t \underline{\underline{\Sigma}}^{-1} \vec{y} = \left(\sum_{l=1}^{N} \underline{\underline{X}}_l^t \underline{\underline{\Sigma}}_l^{-1} \underline{\underline{X}}_l \right)^{-1} \sum_{k=1}^{N} \underline{\underline{X}}_k^t \underline{\underline{\Sigma}}_k^{-1} \vec{y}_k \quad (4)$$

Efficient GLS estimation requires an accurate and *wellconditioned* model for the the covariance structure of the errors. When $\underline{\underline{\Delta}}$ is unknown, an estimate, $\hat{\underline{\underline{\Delta}}}$, of $\underline{\underline{\Delta}}$ can be used. In many estimated GLS schemes, the estimates of $\vec{\beta}$ and $\underline{\underline{\Delta}}$ are iterated. To simplify these expressions we note that Rao's equality (Ref. [10], p. 33) generalizes to

$$(\sigma^2 \underline{\underline{I}} + \underline{\underline{X}} \underline{\underline{\Delta}} \underline{\underline{X}}^t)^{-1} = \sigma^{-2} [\underline{\underline{I}} - \underline{\underline{X}} \underline{\underline{E}} \underline{\underline{X}}^t] + \underline{\underline{X}} \underline{\underline{E}} (\sigma^2 \underline{\underline{E}} + \underline{\underline{\Delta}}_k)^- \underline{\underline{E}} \underline{\underline{X}}^t \quad (5)$$

where $^-$ denotes the Moore-Penrose generalized inverse, $\underline{\underline{E}} = (\underline{\underline{X}}^t \underline{\underline{X}})^-$ and $\underline{\underline{\Delta}}_k$ is the normalized projection of $\underline{\underline{\Delta}}$ onto the row space of $\underline{\underline{X}}_k$, $\underline{\underline{\Delta}}_k = P_{x_k} \underline{\underline{\Delta}} P_{x_k}$.

$$\hat{\vec{\beta}} = \left(\sum_{k=1}^{N} (\sigma_k^2 \underline{\underline{E}}_k + \underline{\underline{\Delta}}_k)^- \right)^{-1} \sum_{l=1}^{N} (\sigma_l^2 \underline{\underline{E}}_l + \underline{\underline{\Delta}}_l)^- \hat{\vec{\beta}}_l = \sum_{l=1}^{N} \underline{\underline{W}}_l \hat{\vec{\beta}}_l \quad (6)$$

where the weighting matrix, $\underline{\underline{W}}_l$ is defined as

$$\underline{\underline{W}}_l = \left(\sum_{k=1}^{N} (\sigma_k^2 \underline{\underline{E}}_k + \underline{\underline{\Delta}}_k)^- \right)^{-1} (\sigma_l^2 \underline{\underline{E}}_l + \underline{\underline{\Delta}}_l)^- \quad (7).$$

This formula is a slight generalization of Swamy's[1] results by allowing the individual covariance matrices $\underline{\underline{X}}_j^t \underline{\underline{X}}_j$ to be singular.

The variance in our estimate of the mean regression vector, $\vec{\beta}$ is given by

$$\underline{\underline{\Omega}} = \left(\sum_{k=1}^{N} (\sigma_k^2 \underline{\underline{E}}_k + \underline{\underline{\Delta}}_k)^- \right)^{-1} \quad (8).$$

We refer the reader to several recent articles[4-5] for a thorough discussion of the asymptotic sample properties of this estimator and the $\vec{\overline{\beta}}$ estimator.

We now concentrate on estimating the random coefficient matrix $\underline{\underline{\Delta}}$ from the residuals, $\hat{\vec{\beta}}_k - \hat{\vec{\beta}}$. There are two competing methodologies for estimating $\underline{\underline{\Delta}}$. The first is to use a simple unbiased estimator introduced by Swamy [1-5]. The second is to use the maximum likelihood (ML) estimator. The ML estimator is asymptotically efficient, but can only be estimated numerically. The ML estimator primarily differs from the Swamy's unbiased estimator by a matrix weighting. In this section, we also present several new estimators which differ from the standard estimates by treating the uncertainty in $\hat{\vec{\beta}}$ explicitly. In particular, we give bias corrections for the ML estimator.

When the individual design matrices, $\underline{\underline{X}}_k^t \underline{\underline{X}}_k$, are degenerate, the differing number of degrees of freedom in the various parameter directions must be accounted for [6,11]. To avoid this difficult problem, we now assume each $\underline{\underline{X}}_k^t \underline{\underline{X}}_k$ is nondegenerate. All of the methods which we discuss below have the difficulty that the estimate, $\hat{\underline{\underline{\Delta}}}$, can have negative eigenvalues. The standard solution to this problem is to set the negative eigenvalues to zero using the singular value decomposition. This procedure results in a positive bias in $\hat{\underline{\underline{\Delta}}}$.

Swamy's unbiased estimator can be summarised as follows. Swamy uses the ordinary least squares (OLS) estimate for the variance of $\hat{\vec{\beta}}_k$. He estimates the RC variance matrix, $\underline{\underline{\Delta}}$, as the difference between the empirical variance of the $\hat{\vec{\beta}}_k$ and the within individual variance. Thus Swamy's esti-

mator for the within individual variance is

$$\hat{\sigma}_k^2 = \frac{\vec{y}_k{}^t P_{\perp_k} \vec{y}_k}{n_k - m} \tag{9}.$$

$P_{\perp,k}$ is the projection perpendicular to the column space of $\underline{\underline{X}}_k$, $P_{\perp,k} = I_k - \underline{\underline{X}}_k \underline{\underline{E}}_k \underline{\underline{X}}_k^t$.

The empirical covariance of the regression coefficients is

$$\underline{\underline{S}}_b = \sum_{k=1}^{N} (\hat{\vec{\beta}}_k - \hat{\bar{\vec{\beta}}})(\hat{\vec{\beta}}_k - \hat{\bar{\vec{\beta}}})^t \tag{10}$$

where $\hat{\bar{\vec{\beta}}}$ is the simple mean of the $\hat{\vec{\beta}}_k$, not the weighted mean of Eq. 6 which corresponds to the GLS estimate of $\hat{\vec{\beta}}$.

Swamy's estimated dispersion matrix, $\hat{\underline{\underline{\Delta}}}$, is the difference between the empirical covariance matrix, $\underline{\underline{S}}_b$ and the theoretically predicted value, calculated under the assumption of identical regression coefficients:

$$\hat{\underline{\underline{\Delta}}} = \frac{1}{N-1} \underline{\underline{S}}_b - \frac{1}{N} \sum_{k=1}^{N} \hat{\sigma}_k^2 (\underline{\underline{X}}_k^t \underline{\underline{X}}_k)^{-1} \tag{11}$$

As an alternative, we now present a new unbiased estimator of $\underline{\underline{\Delta}}$ based on the GLS estimator of $\vec{\beta}$. From the representation of $\hat{\vec{\beta}}$ as the weighted sum of the individual $\hat{\vec{\beta}}_k$ Eq. (6), we calculate the variance of $\delta\hat{\vec{\beta}}_k$:

$$E\left((\hat{\vec{\beta}}_k - \hat{\bar{\vec{\beta}}})(\hat{\vec{\beta}}_k - \hat{\bar{\vec{\beta}}})^t\right) = (\underline{\underline{\Delta}} + \sigma_k^2 \underline{\underline{E}}_k - \underline{\underline{\Omega}}) \tag{12}$$

From Eq. (12), an alternative estimator for $\hat{\underline{\underline{\Delta}}}$ follows from

$$\sum_{k=1}^{N} (\hat{\vec{\beta}}_k - \hat{\bar{\vec{\beta}}})(\hat{\vec{\beta}}_k - \hat{\bar{\vec{\beta}}})^t = N(\hat{\underline{\underline{\Delta}}} - \underline{\underline{\Omega}}) + \sum_{k=1}^{N} \hat{\sigma}_k^2 (\underline{\underline{X}}_k^t \underline{\underline{X}}_k)^{-1} \tag{13}$$

Using Eq. (13), $\hat{\underline{\underline{\Delta}}}$ is iteratively calculated using the previous value of $\hat{\underline{\underline{\Delta}}}$ in $\underline{\underline{\Omega}}$. The estimations of $\vec{\beta}$, and $\underline{\underline{\Delta}}$ are performed iteratively as well. We believe this new estimator will be more efficient since we use a more efficient estimate of $\vec{\beta}$. We note that this modification is a $1/N$ effect.

The ML functional for the RC model[3] reduces to

$$\sum_{k=1}^{N} ln\,|\sigma_k| + ln\left|(\sigma_k^2 \underline{\underline{E}}_k + \underline{\underline{\Delta}})^{-}\right| + (n_k - m_k)\hat{\sigma}_k^2/\sigma_k^2 + \left((\hat{\vec{\beta}}_k - \hat{\vec{\beta}})^t (\sigma_k^2 \underline{\underline{E}}_k + \underline{\underline{\Delta}})^{-} (\hat{\vec{\beta}}_k - \hat{\vec{\beta}})\right) \tag{14}.$$

where the constant term has been ignored. Following [3], minimizing the ML functional with respect to $\underline{\underline{\Delta}}$ yields:

$$\sum_{k=1}^{N} \left((\sigma_k^2 \underline{\underline{E}}_k + \underline{\underline{\Delta}})^{-1} (\hat{\vec{\beta}}_k - \hat{\vec{\beta}})(\hat{\vec{\beta}}_k - \hat{\vec{\beta}})^t (\sigma_k^2 \underline{\underline{E}}_k + \underline{\underline{\Delta}})^{-1} - (\sigma_k^2 \underline{\underline{E}}_k + \underline{\underline{\Delta}})^{-1}\right) = 0 \tag{15}.$$

We note that the ML estimates for σ_k^2 are considerably more complicated than the OLS estimate of Eq. 9. Hybrid estimation schemes, which utilise Eqs. 9 and 15, are of considerable practicle interest.

Eq. 15 can reexpressed in terms of the matrix weighting of Eq. 6:

$$\sum_{k=1}^{N} \underline{\underline{W}}_k \left((\hat{\vec{\beta}}_k - \hat{\vec{\beta}})(\hat{\vec{\beta}}_k - \hat{\vec{\beta}})^t - (\sigma_k^2 \underline{\underline{E}}_k + \underline{\underline{\Delta}})\right) \underline{\underline{W}}_k^t = 0 \tag{16}.$$

Thus the ML estimate of $\underline{\underline{\Delta}}$ is nearly a variance weighted modification of Swamy's estimate. Our bias corrected ML estimate satisfies

$$\sum_{k=1}^{N} \underline{\underline{W}}_k \left((\hat{\vec{\beta}}_k - \hat{\vec{\beta}})(\hat{\vec{\beta}}_k - \hat{\vec{\beta}})^t - (\sigma_k^2 \underline{\underline{E}}_k + \underline{\underline{\Delta}} - \underline{\underline{\Omega}})\right) \underline{\underline{W}}_k^t = 0 \tag{17}.$$

We note that if the between individual variation is much larger than the within individual variance we can expand Eq. 15 in powers of $\sigma_k^2 \underline{\underline{\Delta}}^{-1} \underline{\underline{E}}_k$. To second order, the expansion yields the Swamy estimator of Eq. 11 with N-1 replaced by N. In the opposite limit, when $\underline{\underline{\Delta}} << \sigma_k^2 \underline{\underline{E}}_k$, the expansion yields the weighted equation

$$\sum_{k=1}^{N} (\sigma_k^2 \underline{\underline{E}}_k)^{-1} \left((\hat{\vec{\beta}}_k - \hat{\vec{\beta}})(\hat{\vec{\beta}}_k - \hat{\vec{\beta}})^t - (\sigma_k^2 \underline{\underline{E}}_k + \underline{\underline{\Delta}}) \right) (\sigma_k^2 \underline{\underline{E}}_k)^{-1} = 0 \qquad (18).$$

III. Extensions of Random Coefficient Models to Include Between Covariates

We now consider the case where there are a number of between covariates, $\vec{x}_0$, i.e. variables which are fixed in each individual. We partition the m vector of covariates, $\vec{x}$, into $(\vec{x}_0^t, \vec{x}_1^t)^t$ where $\vec{x}_0$ is a m_0 vector of between covariates and $\vec{x}_1$ is a m_1 vector of within covariates.

When there is only one between covariate, such as the intercept, its value is uniquely determined in each individual and the standard random coefficient model may be applied. If there is more than one between covariate, the values of the between regression coefficients, $\vec{\beta}_0$ are indeterminable within a given individual. In this case, only $\vec{\beta}_1$ and $\vec{\beta}_0 \cdot \vec{x}_{0_k}$ can be determined with a single individual. We note that the GLS estimator of $\vec{\beta}_k$, obtained with the Moore-Penrose inverse, simplifies to

$$\hat{\vec{\beta}}_k = \underline{\underline{E}}_k \underline{\underline{X}}_k^t \vec{y}_k = \begin{pmatrix} c_k \vec{\overline{x}}_{0,k} \\ \hat{\vec{\beta}}_{1,k} \end{pmatrix}$$

where $c_k = (\overline{y}_k - \hat{\vec{\beta}}_{1,k} \cdot \vec{\overline{x}}_{1,k})/|\vec{\overline{x}}_{0,k}|^2$

Since each individual provides an estimate of a different projection of $\vec{\beta}_0$, the entire $\vec{\beta}_0$ can be estimated by combining the various individual estimates of the projections of $\vec{\beta}_0$. We assume that the various individual $\vec{x}_{o_k}$ span the space $\mathbf{X_o}$ and that as the number of individuals increases, every direction in $\mathbf{X_o}$ space has a finite density. However, the distribution of $\vec{x}_{o_k}$ should not be considered homogenous. In many problems of practicle interest, the distribution of $\vec{x}_{o_k}$ values is notably anisotropic with one or two parameter directions accounting for most of the variation. In addition,

the between individual variation in the mean values of $\vec{x}_1$ may be strongly correlated with $\vec{x}_o$. In analyzing different fusion experiments, the author[6-8] found that many covariates scaled with device size.

$\vec{\beta}_0$ may be estimated by a matrix weighting of the estimates of the various projections from each individual using GLS regression on all individuals simultaneously, *provided that the random coefficient matrix* $\underline{\underline{\Delta}}$, *is known. The problem which we address is the determination/modeling of the dispersion matrix,* $\underline{\underline{\Delta}}$. The submatrix of the covariance of $\vec{\beta}_1$, $\underline{\underline{\Delta}}_1$ can be estimated using Eqs. 10-17 with the within variables only. *To complete the specification of* $\underline{\underline{\Delta}}$, *the* $m_0 x m_0$ *dispersion matrix of the between covariates,* $\underline{\underline{\Delta}}_1$ *and the* $m_0 x m_1$ *cross covariance submatrix* $\underline{\underline{\Delta}}_{0,1}$ *must be specified.*

We now discuss the number of different individuals necessary to apply the RC model. In the standard case of Sec. II, we estimate the m parameters of $\vec{\beta}$ plus the $\frac{m(m+1)}{2}$ free parameters of $\underline{\underline{\Delta}}$. Thus the standard rule of statistics is that the number of determinable components of $\hat{\vec{\beta}}_k$, mN, should be at least several times larger than $m + \frac{m(m+1)}{2}$. When $\hat{\underline{\underline{\Delta}}} + \sigma_k^2 \underline{\underline{E}}_k$ is illconditioned, substantially more individuals may be necessary.

In our degenerate case when more than one covariate is fixed in each individual, we have only one projection, $\vec{\beta}_0 \cdot \vec{x}_{0_k}$ for each individual. Thus a total of $(m_1 + 1)N$ determinable components of $\hat{\vec{\beta}}_k$ are available to estimate $m_0 + m_1 + \frac{(m_0+m1)(m_0+m_1+1)}{2}$ free parameters. The actual situation is worse since the N projections, $\vec{\beta}_0 \cdot \vec{x}_{0_k}$, need to be used to estimate the m_o parameters of $\vec{\beta}_o$ plus the $\frac{m_0(m_0+1)}{2}$ parameters of the covariance matrix, $\underline{\underline{\Delta}}_o$, of $\vec{\beta}_o$, plus the $m_0 m_1$ free parameters of the cross covariance of $\vec{\beta}_o$ with

$\vec{\beta}_1$, $\underline{\underline{\Delta}}_{0,1}$. We assume that the number of individuals, N, is large enough to estimate $\vec{\beta}_1$ and $\underline{\underline{\Delta}}_1$, i.e. $Nm_1 >> m_1 + \frac{m_1(m_1+1)}{2}$.

When the number of individuals, N, is sufficiently large, $N >> m_0 + \frac{m_0(m_0+1)}{2} + m_0 m_1$, Δ_0 and $\Delta_{0,1}$ may be completely estimated from the various projections $\vec{\beta}_0 \cdot \vec{x}_{0_k}$ using a modified version of Eqs. 11-17. Naturally all matrices and vectors in eqs. 11-17 must be projected on the column space of $\underline{\underline{X}}_k$. When Eq. 13 or Eq. 17 is used to correct for the uncertainty in the estimate of $\hat{\vec{\beta}}$, our estimates of $\underline{\underline{\Delta}}$ remain unbiased. In contrast, the simple Swamy estimator of Eq.11 generalises to several linear algebraic equations to account for the varying number of degrees of freedom[6]. When the distribution of $\vec{\bar{x}}$ is anisotropic in $\mathbf{X}$ space or $\hat{\underline{\underline{\Delta}}} + \sigma_k^2 \underline{\underline{E}}$ is illconditioned, substantially more individuals are necessary.

We now consider cases where the number of individuals is only slightly larger than m_o, the number of regression parameters which can only be determined by interindividual comparisons. Thus determining $\underline{\underline{\Delta}}_0$ and $\underline{\underline{\Delta}}_{0,1}$ selfconsistently is illconditioned. Instead we must *prescribe a simple functional form for* $\underline{\underline{\Delta}}_0$ *and* $\underline{\underline{\Delta}}_{0,1}$ *and estimate only* $\vec{\beta}$ *and* $\underline{\underline{\Delta}}_1$.

For simplicity, we assume that the between individual covariates include an overall constant and that the overall constant is indexed as the first $\vec{x}_o$ component. As an introduction, we begin with the simple model where $\underline{\underline{\Delta}}_0 \equiv \sigma_o^2 \delta_{1,1}$ and $\underline{\underline{\Delta}}_{0,1} \equiv 0$. $\sigma_o^2 \delta_{1,1}$ corresponds to a random variance of the overall constant. In this mixed model, the absolute constant is random with variance σ_o^2, the parametric dependencies of $\vec{\beta}_o$ are fixed and $\vec{\beta}_1$ is distributed about $\vec{\bar{\beta}}_1$ with covariance matrix Δ_1. Thus the total RC matrix

is

$$\underline{\underline{\Delta}} \equiv \begin{pmatrix} \sigma_o^2 \delta_{1,1} & 0 \\ 0 & \underline{\underline{\Delta}}_1 \end{pmatrix} \tag{19}$$

An important difficulty with this mixed model is that the *parameter and error estimates depend on the choice of within variables.* The within plasma variables, $\vec{x}_1$, can be reparametrised using $\vec{x}_0$, $\vec{x}_{1,A} = \vec{x}_1 + \underline{\underline{A}}\vec{x}_0$ where $\underline{\underline{A}}$ is an arbitrary $m_1 x m_0$ matrix. From a purely statistical point of view, it is impossible to distinguish whether $\vec{x}_1$ or $\vec{x}_{1,A}$ is being varied in a single individual. The models for the errors are not equivalent since we have imposed an arbitrary requirement that the submatrices, $\underline{\underline{\Delta}}_0$ and $\underline{\underline{\Delta}}_{0,1}$ of the random coefficient matrix $\underline{\underline{\Delta}}$ be zero. Thus transforming the within individual variables to $\vec{x}_{1,A}$ changes the model.

In fact, for any linear transformation, $\underline{\underline{A}}$, we can specify a completion of the dispersion matrix by

$$\underline{\underline{\Delta}}_A \equiv \begin{pmatrix} \sigma_o^2 \delta_{o,o} + \underline{\underline{A}}^t \underline{\underline{\Delta}}_1 \underline{\underline{A}} & \underline{\underline{A}}^t \underline{\underline{\Delta}}_1 \\ \underline{\underline{\Delta}}_1 \underline{\underline{A}} & \underline{\underline{\Delta}}_1 \end{pmatrix} \tag{20}.$$

We note that $\vec{\vec{x}}^t \underline{\underline{\Delta}}_A \vec{\vec{x}}$ reduces to $\sigma_o^2 + (\vec{\vec{x}}_1 + \underline{\underline{A}}\vec{\vec{x}}_0)^t \underline{\underline{\Delta}}_1 (\vec{\vec{x}}_1 + \underline{\underline{A}}\vec{\vec{x}}_o)$. $\underline{\underline{\Delta}}_A$ has rank $m_1 + 1$ and the nullspace is spanned by the last $m_o - 1$ columns of$(\underline{\underline{I}}_{m_o}, -\underline{\underline{A}}^t)^t$.

The calculation of $\underline{\underline{\Delta}}_1$ involves only the variation of $y_{i,k}$ and $\vec{x}_{i,k}$ about $\overline{y}_k$ and $\vec{\overline{x}}_k$. In forecasting the performance of new individuals, the quantity $Var(\overline{y}_k - \hat{\vec{\beta}}\vec{\overline{x}}_k)$ is of primary interest. When we use the RC model to infer

the variance of $\overline{y}_k$, we find:

$$Var(\overline{y}_k - \hat{\vec{\beta}}\vec{\overline{x}}_k) = \vec{\overline{x}}_k^t(\underline{\underline{\Delta}} + \sigma_k^2 \underline{\underline{E}}_k - \underline{\underline{\Omega}})\vec{\overline{x}}_k \qquad (21).$$

Unfortunately, the actual empirical distribution of $(\overline{y}_k - \hat{\vec{\beta}}\vec{\overline{x}}_k)$ may differ considerably from the predicted values when the model is incorrectly specified. In particular, if $\delta\vec{\beta}_{0,k} \cdot \vec{\overline{x}}_{0,k}$ and $\delta\vec{\beta}_{1,k} \cdot \vec{\overline{x}}_{1,k}$ are strongly anticorrelated, setting the predicted value of $\delta\vec{\beta}_{0,k}$ to zero results in an overly pessimistic predicted value of $Var(\overline{y}_k - \hat{\vec{\beta}}\vec{\overline{x}}_k)$. In applying this covariance structure, we have imposed an artificially strong heteroscedasticity of the error in fitting the mean values of the individuals as given by Eq. (21). Since we are unable to determine $\underline{\underline{\Delta}}_{0,1}$, we have no basis for assuming a parametric dependence on $Var(\overline{y}_k - \hat{\vec{\beta}}\vec{\overline{x}}_k)$. Thus statistical phenomenology instructs us to choose a homoscedastic or nearly homoscedastic model for the errors.

IV. An Invariant Random Coefficient Model

We now present an alternative dispersion matrix $\underline{\underline{\Delta}}$ which is *a) invariant under linear changes of variable, b) significantly more homoscedastic in the predicted uncertainty of the mean values of the individuals.* We restrict our attention to linear RC models of the form $\underline{\underline{\Delta}}_A$ given by Eq.20. Both σ_o and $\underline{\underline{\Delta}}_1$ are to be determined by fitting the data. We assume that the data is insufficient to determine the $m_1 x m_o$ matrix $\underline{\underline{A}}$. Thus we specify $\underline{\underline{A}}$ by minimizing the weighted heteroscedasticity of the predicted values, $\vec{\overline{x}}_k$,

given $\underline{\underline{\Delta}}_1$.

$$min_A \sum_{k=1}^{N} \lambda_k (\vec{\overline{x}}_{1,k} - \underline{\underline{A}}\,\vec{\overline{x}}_{o,k})^t \underline{\underline{\Delta}}_1 (\vec{\overline{x}}_{1,k} - \underline{\underline{A}}\,\vec{\overline{x}}_{o,k}) \tag{22}$$

where λ_k are unspecified weights. We denote the solution of this minimisation problem by $\underline{\underline{L}}$. Since we focus on the mean values of the individuals, the relevant data matrix is $\underline{\underline{\overline{X}}} = (\underline{\underline{\overline{X}}}_o, \underline{\underline{\overline{X}}}_1)$, which $\underline{\underline{\overline{X}}}$ is a $N \times m$ matrix with rows $\vec{\overline{x}}_k$. We define the NxN diagonal matrix $\underline{\underline{\Lambda}}$ to have the weights λ_k as its diagonal elements.

The weighted minimisation of Eq. (22) corresponds to m_1 separate weighted LS regression of each of the x_1 covariates versus the between covariates. In matrix notation, the solution is

$$\underline{\underline{L}}^t = -(\underline{\underline{\overline{X}}}_0^t \underline{\underline{\Lambda}} \underline{\underline{\overline{X}}}_0)^{-1} \underline{\underline{\overline{X}}}_0^t \underline{\underline{\Lambda}} \underline{\underline{\overline{X}}}_1 \tag{23}$$

Thus for R.C. models of the form given by Eq. 20, the solution, $\underline{\underline{A}} \equiv \underline{\underline{L}}$*, is the most homoscedastic in the sense that* $\underline{\underline{A}} \equiv \underline{\underline{L}}$ *minimizes the parametric variance, averaged over all individuals as expressed by Eq. 22.* We note that the solution matrix $\underline{\underline{L}}$ is independent of $\underline{\underline{\Delta}}_1$ and corresponds to a weighted linear regression of $\vec{\overline{x}}_1$ by $\vec{\overline{x}}_o$.

Since the reparametrisation matrix, $\underline{\underline{A}} \equiv \underline{\underline{L}}$, is specified during model selection phase and not the statistical testing phase, the choice of weights, λ_k, is essentially arbitrary. The weights λ_k may be chosen to correspond to some subjective importance of the datapoints or to the uncertainty in the predicted values, $Var(\overline{y}_k - \hat{\vec{\beta}} \cdot \vec{\overline{x}}_k)$. If the measurement error in the

$\vec{\vec{x}}_k$ variables is significant, the relative measurement errors can be used for λ_k^{-1}. A final method to choose the weights is discussed in the next section. If no convincing reason for nonuniform weights exists, $\lambda_k = 1$ is usually preferred.

As demonstrated in appendix A, the estimated uncertainty for a new observation is invariant under linear coordinate transformations: $\vec{x}_{1,A'} = \vec{x}_1 + \underline{\underline{A}}'\vec{x}_0$.

To estimate the RC matrix parameters, $\underline{\underline{\Delta}}_1$ and σ_o, we can either use a simple unbiased Swamy estimator or an asymptotically efficient ML estimator. For the Swamy estimator, we apply Eq. 11 to the within individual regression parameters to estimate $\underline{\underline{\Delta}}_1$. σ_o can then be estimate by summing Eq. 21 over individuals:

$$N\sigma_o^2 = \sum_{k=1}^{N}(\overline{y}_k - \hat{\vec{\beta}}\vec{\vec{x}}_k)^2 - \sum_{k=1}^{N}\vec{x}_{1,L,k}\underline{\underline{\Delta}}_1\vec{x}_{1,L,k} - \sum_{k=1}^{N}\vec{\vec{x}}_k^t(\sigma_k^2\underline{\underline{E}}_k - \underline{\underline{\Omega}})\vec{\vec{x}}_k \quad (24).$$

where $\vec{x}_{1,L} = \vec{x}_1 + \underline{\underline{L}}\vec{x}_0$. The dependence of $\underline{\underline{\Omega}}$ on σ_o can be easily included by solving Eqs. 6, 24 and 11 or 13 iteratively.

To apply the maximum likelihood estimates of Sec. II, we first make the linear change of variables to $\vec{x}_{1,L} = \vec{x}_1 + \underline{\underline{L}}\vec{x}_0$. The $m_1^2 + 1$ equations for $\underline{\underline{\Delta}}_1$ and σ_o are given by the (1,1) and $(m_o + k, m_o + j)$, $1 \leq j, k \leq m_1$ components of Eq. 15. The estimates of $\vec{\beta}$ and $\underline{\underline{\Delta}}_1$ and σ_o are again iterated.

V. Discussion

The choice of covariance structure to model correlated errors is an im-

portant and usually neglected area of statistics. Statistics research on covariance structures concentrates on a) asymptotic distributions of the estimators of $\vec{\beta}$, $\underline{\underline{\Delta}}$, and $y(\vec{x})$ and b) model testing for acceptance of simpler imbedded covariance structures.

We have focused on the case where the residual fit errors probably possess many more parametric dependencies than can be modeled with the given dataset. A pessimist would remark that this is the standard situation. To supplement our lack of data, we have appealed to statistical phenomenology. The general principle which we have used is "Given a class of models which equally well describe the data, choose the most homoscedastic error structure."

Homoscedascity is especially important when the covariates, $\vec{x}$, are correlated. In this case, strongly heteroscedastic models can seriously misrepresent the data. This can occur when the accepted model underestimates the variance of $\overline{y} - \hat{\vec{\beta}} \cdot \vec{x}$ by neglecting the anticorrelations of $\delta\vec{\beta}_0\vec{\overline{x}}_0$ with $\delta\vec{\beta}_1\vec{\overline{x}}_1$. It can also occur when the model assumes large variances of $\vec{\beta}$ in the principal component directions of $\underline{\underline{\overline{X}}}^t\underline{\underline{\overline{X}}}$ which are poorly determined.

For a given dataset with between covariates, other models for the error correlation matrix may be more appropriate than the RC model described in Sec. IV. Our model suffers from the disadvantage that it is not uniformly more homoscedastic than other choices of the matrix $\underline{\underline{A}}$. In determining $\underline{\underline{A}}$ by minimising Eq. 22, we are more homoscedastic on average than other RC models in the same class. However the transformation to $\vec{x}_L$ amplifies the predicted heteroscedastic variance when $\vec{x}^t_{1,L}\underline{\underline{\Delta}}_1\vec{x}^t_{1,L} > \vec{x}^t_1\underline{\underline{\Delta}}_1\vec{x}^t_1$. We can

determine the direction of $\vec{x}$ which has the largest predicted value of the heteroscedastic variance for a given metric, $\underline{\underline{G}}$ on $\mathbf{X}$. Mathematically we maximize $\vec{x}^t_{1,L}\underline{\underline{\Delta}}_1\vec{x}^t_{1,L}$ subject to the constraint: $\vec{x}^t_1\underline{\underline{G}}\vec{x}^t_1 = 1$. The solution is the eigenvector corresponding to the largest eigenvalue of $\underline{\underline{\Delta}} - \sigma_o\delta_{1,1} - \lambda\underline{\underline{G}}$.

The key step in reducing the average predicted heteroscedasticity has been the linear transformation $\vec{x}_1 \leftarrow \vec{x}_1 + \underline{\underline{L}}\vec{x}_o$. This orthogonalises $\vec{x}_o$ and $\vec{x}_1$ on average and thereby reduces problems associated with interdependent covariates. To account for the between individual variation in $\vec{\beta}_1$, we are forced to use a RC structure containing $\underline{\underline{\Delta}}_1$ as a submatrix. Within the class of RC matrices which include $\underline{\underline{\Delta}}_1$ and are represented in $\vec{x}_L$ coordinates, $\underline{\underline{\Delta}}_L$ is the simplest. As the dataset grows, the order of the model for the randomness should be increased until the full RC matrix is reached. The choice of intermediate models will usually depends greatly on the specific nature of the dataset.

In reality, many datasets contain covariates which vary little within an individual. These covariates are almost but not quite between covariates. Since it is unstable to treat these covariates as within variables, one usually forces these covariates to be between covariates. The backroom procedure is to set the values of these covariates to their mean value, $\overline{x}_k$ and then to correct $y_{k,i}$ for the variation $\vec{x}_{k,i} - \overline{x}_k$. Naturally, the correction depends on $\hat{\vec{\beta}}$ and must therefore be performed iteratively.

Since the uncertainty in forecasting $Var(y - \hat{\vec{\beta}}\vec{x})$ is often of primary interest, validation of Eq. 21 is important. The constant variance, σ_o^2, is to determine to insure Eq. 21 is true on average within the database. Ho-

wever the parametric dependencies $Var(y - \hat{\vec{\beta}}\vec{x})$ may be incorrect. Thus the residuals, $(\overline{y}_k - \hat{\vec{\beta}}\vec{\overline{x}}_k)^2 - \vec{\overline{x}}_k^t(\underline{\underline{\Delta}} + \sigma_k^2\underline{\underline{E}}_k - \underline{\underline{\Omega}})\vec{\overline{x}}_k$, should be examined graphically. The weights, λ_k, can sometimes be used to reduce the parametric dependencies of the residuals. In otherwords, our covariance structure is really $\underline{\underline{\Delta}}_L(\underline{\underline{\Lambda}})$ and $\underline{\underline{\Lambda}}$ may be used to achieve a better fit.

Acknowledgment

This work was performed under U.S. Department of Energy Grant No. DE-FG02-86ER-53223.

Appendix A: Proof of Invariance of $\underline{\underline{\Delta}}_L$ under Linear Transformations

We consider a general linear transformation of the form:

$$\vec{x}_0' = \underline{\underline{C}}\,\vec{x}_0 \ , \quad \vec{x}_1' = \underline{\underline{B}}\,\vec{x}_1 + \underline{\underline{A}}\,\vec{x}_0$$

where $\underline{\underline{C}}$ is a nonsingular $m_o x m_o$ matrix, $\underline{\underline{B}}$ is a nonsingular $m_1 x m_1$ matrix, and $\underline{\underline{C}}$ is a general $m_1 x m_o$ matrix.

The transformation is rewritten in terms of the data matrices as

$$\underline{\underline{X}}_0' = \underline{\underline{X}}_0 \underline{\underline{C}}^t \ , \quad \underline{\underline{X}}_1' = \underline{\underline{X}}_1 \underline{\underline{B}}^t + \underline{\underline{X}}_0 \underline{\underline{A}}^t .$$

The regression coefficients transform to

$$\vec{\beta}_1' = \underline{\underline{B}}^{t,-1} \vec{\beta}_1 \ , \quad \vec{\beta}_0' = \underline{\underline{C}}^{t,-1} \vec{\beta}_0 + \underline{\underline{A}}^t \underline{\underline{B}}^{t,-1} \vec{\beta}_1$$

and the dispersion, $\underline{\underline{\Delta}}_1$ transforms to $\underline{\underline{\Delta}}_1' = \underline{\underline{B}}^{t,-1} \underline{\underline{\Delta}}_1 \underline{\underline{B}}^{-1}$.

The crucial regression matrix, $\underline{\underline{L}}$, transforms to

$$\underline{\underline{L}}'' = -\left(\underline{\underline{C}}\,\underline{\overline{X}}_0^t \underline{\underline{\Lambda}}\,\underline{\overline{X}}_0 \underline{\underline{C}}^t\right)^{-1} \underline{\underline{C}}\,\underline{\overline{X}}_0^t \underline{\underline{\Lambda}} \left(\underline{\underline{X}}_1 \underline{\underline{B}}^t + \underline{\underline{X}}_0 \underline{\underline{A}}^t\right) = \underline{\underline{C}}^{t,-1} \left(\underline{\underline{L}}\,\underline{\underline{B}}^t - \underline{\underline{A}}^t\right).$$

Since $\vec{x}_1 + \underline{\underline{L}}\,\vec{x}_0$ transforms to $\underline{\underline{B}}\left(\vec{x}_1 + \underline{\underline{L}}\,\vec{x}_0\right)$, the variance, $\vec{x}^t \underline{\underline{\Delta}} \vec{x}$ is invariant.

Appendix B: Weighted Ridge Regression

The generalized least squares estimator yields the estimate of $\vec{\beta}$ with the minimum variance within the class of linear unbiased estimators. Ridge regression [13] is a widely used technique to further reduce the variance in the estimate of $\vec{\beta}$ at the cost of introducing bias.

In this appendix, we present a number of the standard results of Hoerl and Kennard for the general case of an arbitrary covariance matrix $\underline{\underline{\Sigma}}$ and ridge matrix $\underline{\underline{P}}$. We are most interested in the case where $\underline{\underline{\Sigma}}$ is generated by a random coefficient model and $\underline{\underline{P}}$ is the projection perpendicular to the space spanned by the overall constant i.e. we do not contract the mean value of our estimates, $\hat{y}_i$.

We let $\underline{\underline{M}}$ denote the generalized design matrix, $X^t\underline{\underline{\Sigma}}^{-1}X$,and U denote $X^t\underline{\underline{\Sigma}}^{-1}Y$. We denote the generalized ridge inverse by $R_\theta = (X^t\underline{\underline{\Sigma}}^{-1}X + \theta\underline{\underline{P}})^{-1}$.

The generalized least square estimate is

$$\hat{\vec{\beta}}_{ls} = (\underline{\underline{X}}^t\underline{\underline{\Sigma}}^{-1}\underline{\underline{X}})^{-1}\underline{\underline{X}}^t\underline{\underline{\Sigma}}^{-1}Y = \beta_0 = R_{\theta=0}U.$$

The generalised ridge estimator is defined as $\hat{\vec{\beta}}_\theta = R_\theta U$ and is related to the least squares estimator by $\hat{\vec{\beta}}_\theta = Z_\theta\hat{\vec{\beta}}$ where $Z_\theta = R_\theta\underline{\underline{M}} = I - \theta R_\theta P$.

To calculate the total mean squared error, we separate the error into a its random and bias components:

$$\hat{\vec{\beta}}_\theta - \vec{\beta} = Z_\theta\hat{\vec{\beta}} - \vec{\beta} = Z_\theta(\hat{\vec{\beta}} - \vec{\beta}) + (Z_\theta\vec{\beta} - \vec{\beta}).$$

The error in the generalised ridge estimator satisfies

$$E(\hat{\vec{\beta}}_\theta - \vec{\beta})^2 = E((\hat{\vec{\beta}} - \vec{\beta})^t Z_\theta^t Z_\theta(\hat{\vec{\beta}} - \vec{\beta})) + (Z_\theta\beta - \beta)^2$$

$$= Trace(\underline{\underline{M}}^{-1} Z_\theta^t Z_\theta) + +\theta^2(\vec{\beta}^t \underline{\underline{P}}\, R_\theta^2 \underline{\underline{P}}\, \vec{\beta})$$

$$= Trace(R_\theta - \theta R_\theta R_\theta P) + +\theta^2(\vec{\beta}^t \underline{\underline{P}}\, R_\theta^2 \underline{\underline{P}}\, \vec{\beta}).$$

The first term is the variance of the estimate and the second term is the bias. Unfortunately, this estimate requires the knowledge of the true value of $\vec{\beta}$. We may replace $\vec{\beta}$ by its ridge estimate, $\hat{\vec{\beta}}_\theta$.

Similarly, the variance of the biased estimate β_θ is

$$E((\hat{\vec{\beta}}_\theta - \vec{\beta})(\hat{\vec{\beta}}_\theta - \vec{\beta})^t) = R_\theta \underline{\underline{M}}\, R_\theta + \theta^2(R_\theta \underline{\underline{P}}\, \vec{\beta}\vec{\beta}^t \underline{\underline{P}}\, R_\theta).$$

Finally the generalised residual sum of squares is

$$SSP = (Y - \underline{\underline{X}}\, \vec{\beta}_\theta)^t \underline{\underline{\Sigma}}^{-1} (Y - \underline{\underline{X}}\, \vec{\beta}_\theta)$$

$$= Y^t \underline{\underline{\Sigma}}^{-1} Y \; - \; U^t(R_\theta + \theta R_\theta P R_\theta)U.$$

For the random coefficient error structure, the first term satisfies

$$Y^t \Sigma^{-1} Y = \sum_k Y_k^t \left(\sigma_k^2 I_k + X_k \hat{\Delta} X_k^t\right)^{-1} Y_k$$

$$= \sum_k \left(\frac{Y_k P_{\perp X_k} Y_k}{\sigma_k^2} \; + \; \vec{\beta}_k^t (\sigma_k^2 E_k + \hat{\Delta})^{-1} \vec{\beta}_k \right).$$

REFERENCES

1. Swamy, P.A.V.B., "Efficient inference in a random coefficient model," *Econometrica*, Vol. 38, No. 2 (1970) p. 311.

2. Swamy, P.A.V.B., "Statistical Inference in Random Coefficient Models," *Lecture Notes in Operations Research and Mathematical Systems*, Vol. 55, Springer Verlag New York (1971)

3. Judge, G.C., Griffiths, W.E., Hill, R.C., Lutkepohl, H. and Lee, T.-C., *Theory and Practice of Econometrics*, J. Wiley and Sons, Inc., New York, 1985.

4. Carter, R.L. and Yang, M.C. "Large Sample Inference in the Random Coefficient Model", *Communications in Statistics- Theory and Practice*, Vol. 15, p. 2507, (1986)

5. Gumpertz, M. and Pantula, S.G., "A Simple Approach to Statistical Inference in the Random Coefficient Model", *American Statistician*, Vol. 43, No. 4, p. 203, (1989)

6. Riedel, K.S., New York University Report MF-118 National Technical Information Service document no. DOE/ER/53223-98 "Advanced Statistics for Tokamak Transport: Colinearity and Tokamak to Tokamak Variation" Copies may be ordered from the National Technical Information Service, Springfield, Virginia 22161. The price is 5.00 plus a 3.00 handling fee. All orders must be prepaid.

7. Riedel, K.S., "Tokamak to tokamak variation and colinearity in scaling laws," *Nuclear Fusion*, Vol. 30, p.755, (1990).

8. Riedel, K.S., and Kaye S.M., "Statistical Analysis of L Mode Energy Confinement", *Nuclear Fusion*, Vol. 30, p.731, (1990).

9. Vonesh, E.F., and Carter, R.L., "Efficient Inference for Random Coefficient Growth Curve Models with Unbalanced Data", *Biometrics*, Vol. 43, p. 617, (1987)

10. Rao, C.R., *Linear Statistical Inference and Its Applications*, J. Wiley and Sons, New York, 1973.

11. Theil, *Principles of Econometrics*, p.322, Wiley, New York, (1971).

12. Barlett, M.S.*Annals of Mathematical Statistics*, vol. 22, p107, (1951)

13. Hoerl, A.E. and Kennard, R.W., *Technometrics*, Vol. 12, p. 55 (1970).