THE PRIOR LIKELIHOOD AND BEST LINEAR UNBIASED PREDICTION IN STOCHASTIC COEFFICIENT LINEAR MODELS

by

Lung-Fei Lee and William E. Griffiths

Discussion Paper No. 79 - 107, January 1979

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, Minnesota 55455

The Prior Likelihood and Best Linear Unbiased Prediction in Stochastic Coefficient Linear Models

by

Lung-Fei Lee and William E. Griffiths*

I. Introduction

In the past decade a great deal of attention has been devoted to variable parameter models and to models which combine the use of time series and cross section data. Balestra and Nerlove (1966), Cooley and Prescott (1976), Hildreth and Houck (1968), Hsiao (1975), Lindley and Smith (1972), Maddala (1971), Swamy (1970, 1971) and Wallace and Hussain (1969), among others, have studied such models. However, most of these studies have tended to concentrate on either estimation of the underlying constant parameters, e.g., Hsiao (1975), or estimation (prediction) of the random components, e.g., Lindley and Smith (1972). There have been a few exceptions, e.g., Swamy (1971). Also, the lack of a sampling theory approach which can handle joint estimation of both constant and random parameters has been regarded as a major deficiency by some Bayesian statisticians e.g., Smith (1973).

In this paper, for error components models and a wide class of random coefficient models, we give a unified approach to the joint estimation of random components and constant underlying parameters. Estimation of random components, the aspect which has received the lesser attention in the past, is useful for predicting future values of the

---

*Lung-Fei Lee, Department of Economics, University of Minnesota, Minneapolis, Minnesota. Financial support received from the National Science Foundation under Grant SOC-78-07304.

William E. Griffiths, Department of Economic Statistics, University of New England, Armidale, NSW, Australia. The paper was written while this author was on study leave at the Department of Economics, University of Minnesota.

We would like to thank C. A. Sims for his valuable comments on an earlier draft. Responsibility for any errors is solely our own.

dependent variable and, for describing the past behavior of a particular random component.

Under the assumption of normality of disturbances we show how the prior likelihood approach, Edwards (1969), can be used to yield "maximum likelihood estimates" of all the parameters. Also, without the normality assumption, these estimates can be regarded as those which result from a "modified mixed estimation procedure". If the variances are assumed known, these procedures lead to an estimator of the constant underlying parameters which is equal to the generalized least squares (GLS) estimator obtained from an integrated equation. This estimator has, of course, its usual properties. The properties of the estimator (predictor) of the realized, but unobservable, random components are not immediately obvious. However this estimator can be shown to be best linear unbiased. With known variances, normality and a diffuse prior on the constant underlying parameters, both estimators are identical to the posterior means of the relevant Bayesian densities.

In section 2, the prior likelihood approach and a modified mixed estimation procedure are introduced. In section 3, we assume all variances are known and discuss estimation of parameters in the error components models and, in section 4, we introduce a general model which includes the Hildreth and Houck (1968), Swamy (1970) and Hsiao (1975) models as special cases. The prior likelihood or modified mixed estimators are given and their best linear unbiased properties demonstrated. The equivalence of these results with some Bayesian estimators with known variances is shown in section 5. The unknown variances cases are considered in section 6. For these cases, we indicate how estimates can be obtained via iterative solution of a set of nonlinear equations and discuss consistency. The difference between this and the Bayesian approach for unknown variances is pointed out.

2. Prior Likelihood and a Modified Mixed Estimation Procedure

The mixed estimation procedure in Theil and Goldberger (1960) incorporates stochastic prior information

$$\underset{q\times 1}{r} = \underset{q\times k}{R}\ \underset{k\times 1}{\beta} + \underset{q\times 1}{v} \tag{2.1}$$

into the linear model

$$\underset{N\times 1}{y} = \underset{N\times k}{X}\ \underset{k\times 1}{\beta} + \underset{N\times 1}{\varepsilon} \tag{2.2}$$

where X is a matrix of exogenous variables with rank k, R is a matrix with rank q, $v\sim(0,\Omega)$, $\varepsilon\sim(0,\Sigma)$ and ε, v are independent.[1] In equations (2.1) and (2.2), β is a vector of constant (fixed) unknown parameters and r is an observed random vector which, in addition to the sample y, contains information about β. The Theil-Goldberger mixed estimation procedure is, assuming Σ and Ω known, to derive an estimator $\hat{\beta}$ of β by minimizing the quadratic form,

$$Q(\beta) = (y-X\beta)'\Sigma^{-1}(y-X\beta) + (r-R\beta)'\Omega^{-1}(r-R\beta). \tag{2.3}$$

This leads to the mixed estimator $\hat{\beta} = (X'\Sigma^{-1}X+R'\Omega^{-1}R)^{-1}(X'\Sigma^{-1}y+R'\Omega^{-1}r)$ which is an unbiased estimator of β and which has a variance matrix, $v(\hat{\beta}) = (X'\Sigma^{-1}X + R'\Omega^{-1}R)^{-1}$ smaller[2] than the variance matrix of $\hat{\beta}_G = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$, the estimator which utilizes sample information in (2.2) only. Thus, for stochastic prior information of the form

---

[1] $v \sim (0,\Omega)$ means that v is a random vector with zero mean and nonsingular covariance matrix Ω.

[2] V_1 is smaller than V_2 means that V_2-V_1 is nonnegative definite.

given in (2.1), the mixed estimation procedure is more efficient, is distribution free and does not involve a Bayesian argument.

However, in many models like (2.2), one often has stochastic prior information on β which cannot be expressed in the form given in (2.1). For example, in a random coefficients model studied by Hildreth and Houck (1968),

$$(2.4) \qquad \underset{1\times 1}{y_i} = \underset{1\times k}{x_i}\ \underset{k\times 1}{\beta_i} + \underset{1\times 1}{\varepsilon_i} \qquad i=1,\ldots,N$$

one has prior information that β_i is random with common unknown mean β and variance Δ. In this case because each β_i is not observable, one cannot express the prior information in the form given in (2.1).

To incorporate such prior information, one needs to generalize the Theil-Goldberger procedure. Consider the linear model with some coefficients β_1 stochastic,

$$(2.5) \qquad \underset{N\times 1}{y} = \underset{N\times k_1}{X_1}\ \underset{k_1\times 1}{\beta_1} + \underset{N\times k_2}{X_2}\ \underset{k_2\times 1}{\beta_2} + \underset{N\times 1}{\varepsilon}$$

$$(2.6) \qquad \underset{k_1\times 1}{\beta_1} = \underset{k_1\times \ell}{A}\ \underset{\ell\times 1}{\Theta} + \underset{k_1\times 1}{\xi}$$

where $\varepsilon \sim (0, \Sigma)$, $\xi \sim (0, \Omega)$, ε, ξ are independent $\ell \leq k_1$, the exogenous variables matrix (X_1, X_2) has rank $k = k_1 + k_3$, and A is a matrix with rank ℓ. With known variances Σ and Ω, to incorporate the prior information (2.6) with sample information (2.5), we suggest minimizing the quadratic form.

$$(2.7) \qquad Q(\beta,\Theta) = (y-X_1\beta_1-X_2\beta_2)'\Sigma^{-1}(y-X_1\beta_1-X_2\beta_2) + (\beta_1-A\Theta)'\Omega^{-1}(\beta_1-A\Theta)$$

with respect to both random and fixed parameters. That is, with respect to β_1, β_2 and Θ.

Thus, to incorporate weak stochastic prior information, this procedure formally generalizes that of Theil and Goldberger and it is also distribution free. It can be justified with the concept of "prior likelihood" introduced in Edwards (1969). He argues that inferences should be based on a likelihood function which combines information from both the "experimental likelihood" and the "prior likelihood". If, in equations (2.5) and (2.6), we assume ε and ξ are multivariate normal then, according to Edward's scheme of inference, the relevant likelihood function for our model is

$$(2.8) \qquad L = \frac{1}{(2\Pi)^{N/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(y-X_1\beta_1-X_2\beta_2)'\Sigma^{-1}(y-X_1\beta_1-X_2\beta_2)\}$$

$$\frac{1}{(2\Pi)^{k/2}|\Omega|^{1/2}} \exp\{-\frac{1}{2}(\beta_1-A\Theta)'\Omega^{-1}(\beta_1-A\Theta)\}$$

where the second term is the prior likelihood expressing the fact that β_1 is regarded as a random sample from a normal population with mean $A\Theta$ and variance Ω. It is clear that the estimates derived from (2.7) will be equivalent to these obtained by maximizing (2.8) with respect to β and Θ.

For a simple random coefficient model with no explanatory variables (a special case of (2.5) and (2.6)), this use of the prior likelihood in the context of maximum likelihood estimation has been suggested by Nelder (1972). However, because the procedure differs from the usual maximum likelihood approach, there is a need to evaluate the properties of the resulting estimators. In the following sections, for some variance component and random coefficient regression models, we derive the estimators which minimize (2.7) and evaluate their properties.

3. Variance Components Models

Variance components models have been studied by Balestra and Nerlove (1966), Wallace and Hussain (1969), Nerlove (1971), Maddala (1971) and others. The model with only individual effects is

$$(3.1)\qquad \underset{1\times 1}{y_{it}} = \underset{1\times k}{x_{it}}\ \underset{k\times 1}{\beta} + \underset{1\times 1}{u_i} + \underset{1\times 1}{\varepsilon_{it}} \qquad i=1,\dots,N;\quad t=1,\dots,T$$

where x_{it} is a vector of exogenous variables without intercept term, $u_i \sim$ i.i.d (α,σ_u^2), $\varepsilon_{it} \sim$ i.i.d. $(0,\sigma^2)$ and u_i, ε_{jt} are independent for all i, $j=1,\dots,N$ and $t=1,\dots,T$. The econometric problem that has been studied extensively is the efficient estimation of the population parameters α and β. However, for prediction purposes, as discussed later, the estimation of the realizations u_i, $i=1,2,\dots,N$, plays an important role which has been overlooked in the literature.

Assuming σ_u^2 and σ^2 are known, the proposed modified mixed estimation procedure is to minimize the quadratic function,

$$Q(\alpha,\beta,u_i) = \frac{1}{\sigma^2}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}(y_{it}-x_{it}\beta-u_i)^2 + \frac{1}{\sigma_u^2}\Sigma_{i=1}^{N}(u_i-\alpha)^2$$

with respect to α, β and all u_i. The estimates are obtained by solving the first order conditions:

$$(3.2)\qquad \Sigma_{i=1}^{N}\Sigma_{t=1}^{T} x_{it}'(y_{it}-x_{it}\tilde{\beta} - \tilde{u}_i) = 0$$

$$(3.3)\qquad \frac{1}{\sigma^2}\Sigma_{t=1}^{T}(y_{it} - x_{it}\tilde{\beta} - \tilde{u}_i) - \frac{1}{\sigma_u^2}(\tilde{u}_i-\tilde{\alpha}) = 0 \qquad i=1,\dots,N$$

$$(3.4)\qquad \frac{1}{\sigma_u^2}\Sigma_{i=1}^{N}(\tilde{u}_i-\tilde{\alpha}) = 0.$$

Equation (3.4) implies

$$\tilde{\alpha} = \frac{1}{N} \Sigma_{i=1}^{N} \tilde{u}_i \, . \qquad (3.5)$$

In turn, equations (3.3) and (3.5) imply

$$\tilde{\alpha} = \bar{y} - \bar{x}\tilde{\beta} \qquad (3.6)$$

where $\bar{y} = \frac{1}{NT} \Sigma_{i=1}^{N} \Sigma_{t=1}^{T} y_{it}$ and $\bar{x} = \frac{1}{NT} \Sigma_{i=1}^{N} \Sigma_{t=1}^{T} x_{it}$ are sample means. With (3.6), it follows from (3.3) that

$$\tilde{u}_i = \frac{1}{T/\sigma^2 + 1/\sigma_u^2} \left(\frac{T}{\sigma^2} (\bar{y}_i - \bar{x}_i \tilde{\beta}) + \frac{1}{\sigma_u^2} (\bar{y} - \bar{x}\tilde{\beta}) \right) \qquad i=1,\ldots,N \qquad (3.7)$$

where $\bar{y}_i = \frac{1}{T} \Sigma_{t=1}^{T} y_{it}$ and $\bar{x}_i = \frac{1}{T} \Sigma_{t=1}^{T} x_{it}$. Substituting $\tilde{u}_i$ into (3.2) and rearranging terms yields

$$\tilde{\beta} = \left(\frac{1}{NT} \Sigma_{i=1}^{N} \Sigma_{t=1}^{T} x_{it}' x_{it} - \frac{T\sigma_u^2}{T\sigma_u^2 + \sigma^2} \frac{1}{N} \Sigma_{i=1}^{N} \bar{x}_i' \bar{x}_i - \frac{\sigma^2}{T\sigma_u^2 + \sigma^2} \bar{x}' \bar{x} \right)^{-1} \cdot \qquad (3.8)$$

$$\left(\frac{1}{NT} \Sigma_{i=1}^{N} \Sigma_{t=1}^{T} x_{it}' y_{it} - \frac{T\sigma_u^2}{T\sigma_u^2 + \sigma^2} \frac{1}{N} \Sigma_{i=1}^{N} \bar{x}_i' \bar{y}_i - \frac{\sigma^2}{T\sigma_u^2 + \sigma^2} \bar{x}' \bar{y} \right).$$

From (3.6) and (3.8) it is clear that the estimator $(\tilde{\alpha}, \tilde{\beta})$ is equal to Aitken's generalized least squares (GLS) estimator derived from the integrated equation $y_{it} = \alpha + x_{it}\beta + v_{it}$ where $v_{it} = (u_i - \alpha) + \varepsilon_{it}$. Thus, for the fixed population parameters, the modified mixed estimation procedure leads to estimators which are best linear unbiased with $\tilde{\beta}$ utilizing the within group and between group information (Maddala (1971)).

The estimator derived in (3.7) is a weighted average of $\hat{u}_i = \bar{y}_i - \bar{x}_i \tilde{\beta}$, an estimator based only on observations on the i^{th} individual, and

$\tilde{\alpha}$, the estimator of the overall mean α. Thus $\tilde{u}_i$ shifts $\hat{u}_i$ towards a common mean. This estimator $\tilde{u}_i$ is intuitively appealing since the smaller the variance σ_u^2, the more likely it is that the u_i's will be close to the mean α. Also, unlike $\hat{u}_i$, it uses all the available information. Equation (3.7) can be written as

$$\tilde{u}_i = \tilde{\alpha} + \frac{T\sigma_u^2}{T\sigma_u^2 + \sigma^2} (\bar{y}_i - \tilde{\alpha} - \bar{x}_i \tilde{\beta}) \tag{3.9}$$

and, in this form, it can be interpreted as being equal to our estimate of the mean plus a proportion of the residual $(\bar{y}_i - \tilde{\alpha} - \bar{x}_i \tilde{\beta})$, allocated to $(u_i - \alpha)$.

Conditional on the realizations u_i, it is clear that $\tilde{u}_i$ is a biased estimator of u_i. Unconditionally, however, it is possible to show that $\tilde{u}_i$ is the best linear unbiased predictor of u_i. The unconditional inference is well justified with respect to repeated sampling from the joint probability distribution of (y_{it}, u_i). In fact, the unconditional inference argument is used to show that the GLS estimator $\tilde{\beta}$ is BLUE.

To demonstrate that $\tilde{u}_i$ is best linear unbiased, we digress for a moment to state, in the more general form of section 2, a vector extension of a result from Goldberger (1962). See also Theil (1971). The "integrated equation" from (2.5) and (2.6) is

$$y = X_1 A \Theta + X_2 \beta_2 + \varepsilon + X_1 \xi \tag{3.10}$$

or

$$y = F\gamma + u \tag{3.11}$$

with $F = (X_1A, X_2)$, $\gamma' = (\Theta', \beta_2')$ and $u = \varepsilon + X_1\xi$. Also, equation (2.6) can be written as

$$(3.12) \qquad \beta_1 = G\gamma + \xi$$

where $G = (A, 0)$ and

$$(3.13) \qquad \text{cov}\binom{u}{\xi} = \begin{pmatrix} \Sigma + X_1\Omega X_1' & X_1\Omega \\ \Omega X_1' & \Omega \end{pmatrix} = \begin{pmatrix} V & W \\ W' & \Omega \end{pmatrix}$$

From equations (3.11), (3.12) and (3.13) and a theorem in Theil (1971, p. 280) the predictor

$$(3.14) \qquad \hat{\beta}_1 = G\hat{\gamma} + W'V^{-1}(y-F\hat{\gamma})$$

with $\hat{\gamma} = (F'V^{-1}F)^{-1}F'V^{-1}y$, is best linear unbiased in the following sense: With respect to repeated sampling from the joint distribution of y and β_1 (or simply, unconditional on β_1), $\hat{\beta}_1$ is unbiased, $E(\hat{\beta}_1 - \beta_1) = 0$, and the prediction error of any predictor of β_1, which is also linear in y and unbiased, has a covariance matrix which exceeds that of $\hat{\beta}_1$ by a positive semi-definite matrix. The covariance matrix of the prediction error is

$$(3.15) \qquad \text{cov}(\hat{\beta}_1-\beta_1) = G(F'V^{-1}F)^{-1}G' + \Omega - W'(V^{-1}-V^{-1}F(F'V^{-1}F)^{-1}F'V^{-1})W$$
$$-G(F'V^{-1}F)^{-1}F'V^{-1}W - W'V^{-1}F(F'V^{-1}F)^{-1}G'$$

Partitioning $\hat{\gamma}$ as $\hat{\gamma}' = (\hat{\Theta}', \hat{\beta}_2')$ and using the definitions of W and V, equation (3.14) can be rewritten as

$$(3.16) \qquad \hat{\beta}_1 = A\hat{\Theta} + \Omega X_1'(\Sigma + X_1\Omega X_1')^{-1}(y - X_1A\hat{\Theta} - X_2\hat{\beta}_2) .$$

One should note that, with respect to probability distribution of y given β_1, i.e., conditional on β_1, $\hat{\beta}_1$ will be a biased estimator.

We shall now put our model in equation (3.1) into the more general form and hence show that $\tilde{u}_i$ of equation (3.7) is BLUE. Let $y = (y_{11},\ldots,y_{1T},\ldots,y_{N1},\ldots,y_{NT})'$, $X_2 = (x'_{11},\ldots,x'_{1T},\ldots,x'_{N1},\ldots,x'_{NT})'$, $X_1 = I_N \otimes \ell_T$, $A = \ell_N = (1,\ldots,1)'$, $\beta_1 = (u_1,\ldots,u_N)'$, $\Theta = \alpha$, $\beta_2 = \beta$, $\Sigma = \sigma^2 I_{NT}$, $\Omega = \sigma_u^2 I_N$. Then (3.16) becomes

$$(3.17) \qquad \tilde{u} = \ell_N\tilde{\alpha} + (I_N \otimes \sigma_u^2\, \ell_T')\, \frac{1}{\sigma^2}\, (I_{NT} - \frac{\sigma_u^2}{\sigma^2 + T\sigma_u^2}\, I_N \otimes \ell_T\ell_T') \cdot$$

$$(y - X\tilde{\beta} - \tilde{\alpha}\ell_{NT})$$

$$= \ell_N\tilde{\alpha} + \frac{T\sigma_u^2}{\sigma^2 + T\sigma_u^2} \cdot (\frac{1}{T}\, I_N \otimes \ell_T\ell_T')(y - X\tilde{\beta} - \tilde{\alpha}\ell_{NT})$$

and this is equivalent to (3.9) written as a vector. The covariance matrix of the prediction error, obtained by making the appropriate substitutions into (3.15) simplifying is given by

$$(3.18) \qquad \operatorname{cov}(\tilde{u}_i - u_i)$$

$$= \frac{\sigma^2}{T\sigma_u^2 + \sigma^2}\, (\sigma_u^2 + \frac{\sigma^2}{NT}) + \frac{\sigma^2}{NT}\, (\frac{T\sigma_u^2}{T\sigma_u^2 + \sigma^2})^2 \bar{x}_i' \cdot$$

$$(\frac{X'X}{NT} - \frac{T\sigma_u^2}{T\sigma_u^2 + \sigma^2}\, \frac{1}{N}\sum_{i=1}^{N} \bar{x}_i'\bar{x}_i)^{-1}\bar{x}_i' \; .$$

Following the same argument as in Goldberger (1962), given a set of exogenous variables x_{iF}, $i=1,\ldots,N$, the predictors

$$(3.19) \qquad \hat{y}_{iF} = x_{iF}\tilde{\beta} + \tilde{u}_i$$

are BLUE predictors for $y_{iF} = x_{iF}\beta + u_i + \varepsilon_{iF}$.

The analysis above can be extended to variance components models with both time and individual effects. This model can be written as

$$(3.20) \qquad \begin{aligned} &\underset{1\times 1}{y_{it}} = \underset{1\times k}{x_{it}} \underset{k\times 1}{\beta} + \underset{1\times 1}{\alpha_i} + \underset{1\times 1}{\gamma_t} + \underset{1\times 1}{\varepsilon_{it}} \\ &\alpha_i = \alpha + u_i \\ &\gamma_t = v_t \end{aligned}$$

where ε_{it}, u_i, v_t are independent with zero means and (known) variances σ^2, σ_u^2 and σ_v^2. The mixed estimation problem is to jointly estimate the fixed parameters (α, β) and the parameters with random status α_i and γ_t. The estimators are found by solving the following first order equations

$$(3.21) \qquad \Sigma_{i=1}^{N}\Sigma_{t=1}^{T} x_{it}'(y_{it} - x_{it}\tilde{\beta} - \tilde{\alpha}_i - \tilde{\gamma}_t) = 0$$

$$(3.22) \qquad \frac{1}{\sigma^2}\Sigma_{t=1}^{T}(y_{it} - x_{it}\tilde{\beta} - \tilde{\alpha}_i - \tilde{\gamma}_t) - \frac{1}{\sigma_u^2}(\tilde{\alpha}_i - \tilde{\alpha}) = 0 \qquad i=1,\ldots,N$$

$$(3.23) \qquad \frac{1}{\sigma^2}\Sigma_{i=1}^{N}(y_{it} - x_{it}\tilde{\beta} - \tilde{\alpha}_i - \tilde{\gamma}_t) - \frac{1}{\sigma_v^2}\tilde{\gamma}_t = 0 \qquad t=1,\ldots,T$$

$$(3.24) \qquad \Sigma_{i=1}^{N}(\tilde{\alpha}_i - \tilde{\alpha}) = 0 \,.$$

Without loss of generality, one can assume $\bar{x} = 0$. With elementary manipulations, one has

$$(3.25) \qquad \tilde{\alpha} = \frac{1}{N} \Sigma_{i=1}^{N} \tilde{\alpha}_i = \bar{y}$$

$$(3.26) \qquad \tilde{\alpha}_i = \frac{1}{T\sigma_u^2 + \sigma^2} [T\sigma_u^2(\bar{y}_i - \bar{x}_i \tilde{\beta}) + \sigma^2 \bar{y}] \qquad i=1,\ldots,N$$

$$(3.27) \qquad \tilde{\gamma}_t = \frac{N\sigma_v^2}{N\sigma_v^2 + \sigma^2} (\bar{y}_{\cdot t} - \bar{x}_{\cdot t}\tilde{\beta} - \tilde{\alpha}) \qquad t = 1,\ldots,T,$$

where $\bar{x}_{\cdot t} = \frac{1}{N} \Sigma_{i=1}^{N} x_{it}$, $\bar{y}_{\cdot t} = \frac{1}{N}\Sigma_{i=1}^{N} y_{it}$. With equations (3.25), (3.26) and (3.27), equation (3.21) becomes

$$(3.28) \qquad \left(\frac{1}{NT} \Sigma_{i=1}^{N}\Sigma_{t=1}^{T} x_{it}' x_{it} - \frac{T\sigma_u^2}{T\sigma_u^2 + \sigma^2} \frac{1}{N}\Sigma_{i=1}^{N} \bar{x}_i' \bar{x}_i - \frac{N\sigma_v^2}{N\sigma_v^2 + \sigma^2} \frac{1}{T}\Sigma_{t=1}^{T} \bar{x}_{\cdot t}' \bar{x}_{\cdot t}\right)\tilde{\beta}$$

$$= \frac{1}{NT} \Sigma_{i=1}^{N}\Sigma_{t=1}^{T} x_{it}' y_{it} - \frac{T\sigma_u^2}{T\sigma_u^2 + \sigma^2} \frac{1}{N}\Sigma_{i=1}^{N} \bar{x}_i' \bar{y}_i - \frac{N\sigma_v^2}{N\sigma_v^2 + \sigma^2} \frac{1}{T}\Sigma_{t=1}^{T} \bar{x}_{\cdot t}' \bar{y}_{\cdot t} \; .$$

It is immediately obvious that $\tilde{\beta}$ solved from the normal equation in (3.28) is exactly the GLS estimator derived in Wallace and Hussain (1969) and Nerlove (1971). Also, one can show that $\tilde{\alpha}_i$ and $\tilde{\gamma}_t$ in (3.26) and (3.27) are best linear unbiased estimators of the random variables α_i and γ_t. For predicting y_{iF} given x_{iF}, as mentioned above in equation (3.19), $\tilde{\alpha}_i$ together with $\tilde{\beta}$ provides the best linear unbiased predictor.

4. Random Coefficient Models

In the econometric literature, there is a wide class of models in which the coefficient vector is stochastic. See, for example, Hildreth and Houck (1968), Swamy (1970), Hsiao (1975), Cooley and Prescott (1973) and Rosenberg (1973a). Instead of analyzing each model individually we shall give some results for a general model and then show how they relate to one of the specific models, namely, that of Swamy (1970).

Consider the following general model

$$(4.1) \qquad \underset{m\times 1}{y} = \underset{m\times k}{X}\ \underset{k\times 1}{\beta} + \underset{m\times 1}{\varepsilon}$$

where X is a matrix of exogenous variable, $\varepsilon \sim (0, C_1)$ with C_1 a known nonsingular variance matrix and β is stochastic related to a vector of unknown fixed parameters Θ by

$$(4.2) \qquad \underset{k\times 1}{\beta} = \underset{k\times \ell}{A}\ \underset{\ell\times 1}{\Theta} + \underset{k\times 1}{\xi}, \qquad \xi \sim (0, C_2),$$

where C_2 is a known nonsingular variance matrix and XA has rank ℓ.

The mixed estimation procedure which combines the sample information in (4.1) with the prior information in (4.2) is defined as

$$\min_{\beta,\Theta} (y-X\beta)'C_1^{-1}(y-X\beta) + (\beta-A\Theta)'C_2^{-1}(\beta-A\Theta).$$

The first order conditions are

$$(4.3) \qquad -X'C_1^{-1}(y-X\tilde{\beta}) + C_2^{-1}(\tilde{\beta}-A\tilde{\Theta}) = 0$$

$$(4.4) \qquad -A'C_2^{-1}(\tilde{\beta}-A\tilde{\Theta}) = 0$$

which imply

$$(4.5) \qquad A'C_2^{-1}A\tilde{\Theta} = A'C_2^{-1}\tilde{\beta}$$

$$(4.6) \qquad (X'C_1^{-1}X + C_2^{-1})\tilde{\beta} = X'C_1^{-1}y + C_2^{-1}A\tilde{\Theta}$$

From (4.6), $\tilde{\beta}$ is a weighted average of two estimators

$$(4.7) \qquad \tilde{\beta} = (X'C_1^{-1}X + C_2^{-1})^{-1}(X'C_1^{-1}X\hat{\beta}_L + C_2^{-1}\hat{\beta}_*)$$

where $X'C_1^{-1}X\hat{\beta}_L = x'C_1^{-1}y$ is the normal equation of β utilizing only the sample information and $\hat{\beta}_* = A\tilde{\Theta}$ is an estimator of the prior mean $A\Theta$. Thus the estimator $\tilde{\beta}$ shifts the sample estimator $\hat{\beta}_L$ towards the prior mean estimator. From (4.5), $\tilde{\Theta}$ is a regression estimate from (4.2) with $\tilde{\beta}$ as regressand,

$$(4.8) \qquad \tilde{\Theta} = (AC_2^{-1}A)^{-1}A'C_2^{-1}\tilde{\beta}$$

More explicitly, $\tilde{\Theta}$ and $\tilde{\beta}$ can be derived from the linear equations (4.5) and (4.6). This yields

$$(4.9) \qquad \tilde{\beta} = (X'C_1^{-1}X + C_2^{-1})^{-1}(X'C_1^{-1}y + C_2^{-1}A\tilde{\Theta})$$

and

$$(4.10) \qquad \tilde{\Theta} = (A'C_2^{-1}A - A'C_2^{-1}(X'C_1^{-1}X+C_2^{-1})^{-1}C_2^{-1}A)^{-1}A'C_2^{-1}(X'C_1^{-1}X+C_2^{-1})^{-1}X'C_1^{-1}y.$$

To evaluate the statistical properties of $\tilde{\Theta}$ and $\tilde{\beta}$, let us first compare $\tilde{\Theta}$ with the GLS estimator $\hat{\Theta}$ derived from the integrated equation

(4.11) $$y = XA\Theta + v$$

where $v = \varepsilon + X\xi$. The GLS estimator $\hat{\Theta}$ is

(4.12) $$\hat{\Theta} = (A'X'(X_1 + XC_2X')^{-1}XA)^{-1}A'X'(C_1+XC_2X')^{-1}y \ .$$

The equivalence of $\tilde{\Theta}$ and $\hat{\Theta}$ is not obvious. However they can be shown to be the same as follows.

The matrix identity $(C_2^{-1} + X'C_1^{-1}X)^{-1} = C_2 - C_2X'(C_1+XC_2X')^{-1}XC_2$ implies,

$$A'X'(C_1 + XC_2X')^{-1}XA$$

$$= A'(C_2^{-1} - C_2^{-1}(C_2^{-1}+X'C_1^{-1}X)^{-1}C_2^{-1})A$$

and

$$A'C_2^{-1}(X'C_1^{-1}X + C_2^{-1})^{-1}X'C_1^{-1}$$

$$= A'C_2^{-1}(C_2 - C_2X'(C_1+XC_2X')^{-1}XC_2)X'C_1^{-1}$$

$$= A'X'(I-(C_1+XC_2X')^{-1}XC_2X')C_1^{-1}$$

$$= A'X'(C_1 + XC_2X')^{-1}$$

Therefore $\tilde{\Theta} = \hat{\Theta}$, i.e., the modified mixed estimation procedure gives the BLUE for Θ.

To evaluate $\tilde{\beta}$ in (4.9), one can use the identity $(X'C_1^{-1}X + C_2^{-1})^{-1} = C_2 - C_2X'(C_1+XC_2X')^{-1}XC_2$, to note that $(X'C_1^{-1}X + C_2^{-1})^{-1}X'C_1^{-1} = C_2X'(C_1 + XC_2X')^{-1}$ and $(X'C_1^{-1}X+C_2^{-1})^{-1}C_2^{-1}A = A - C_2X'(C_1+XC_2X')^{-1}XA$. Hence $\tilde{\beta}$ can be rewritten as

$$\begin{aligned}\tilde{\beta} &= C_2X'(C_1+XC_2X')^{-1}y + (A-C_2X'(C_1+XC_2X')^{-1}XA)\tilde{\Theta} \\ &= A\tilde{\Theta} + C_2X'(C_1 + XC_2X')^{-1}(y-XA\tilde{\Theta}).\end{aligned}$$

This is precisely of the form given in (3.16) and thus $\tilde{\beta}$ is the best linear unbiased estimator (predictor) of β. The covariance matrix of the prediction error is (equation 3.15).

$$\begin{aligned}(4.13)\qquad \text{cov}(\tilde{\beta}-\beta) &= A(A'X'(C_1+XC_2X')^{-1}XA)^{-1}A' + C_2-C_2X'(C_1+XC_2X')^{-1}XC_2 \\ &+ C_2X'(C_1+XC_2X')^{-1}XA(A'X'(C_1+XC_2X')^{-1}XA)^{-1}A'X'(C_1+XC_2X')^{-1}XC_2 \\ &- A(A'X'(C_1+XC_2X')^{-1}XA)^{-1}A'X'(C_1+XC_2X')^{-1}XC_2 \\ &- C_2X'(C_1+XC_2X')^{-1}XA(A'X'(C_1+XC_2X')^{-1}XA)^{-1}A'.\end{aligned}$$

When $\Theta = \bar{\Theta}$ is a known vector, $\tilde{\beta}$ derived from the modified mixed estimation procedure is

$$\begin{aligned}(4.14)\qquad \tilde{\beta} &= (X'C_1^{-1}X + C_2^{-1})^{-1}(X'C_1^{-1}y + C_2^{-1}A\bar{\Theta}) \\ &= A\bar{\Theta} + C_2X'(C_1 + XC_2X')^{-1}(y-XA\bar{\Theta}) .\end{aligned}$$

Unconditional on β, it follows from Chipman (1964) that, from the class of linear estimators, $\tilde{\beta}$ in (4.14) is the minimum mean square error estimator. Thus, for the special case of Θ being known, our modified mixed estimation procedure is formally equivalent to Chipman's approach.

Let us now illustrate the implications for the Swamy-type random coefficient model,

$$(4.15)\qquad \begin{aligned} y_{it} &= x_{it}\beta_i + \varepsilon_{it} \\ \beta_i &= \Theta + v_i \qquad\qquad i=1,\ldots,N;\ t=1,\ldots,T, \end{aligned}$$

where $v = (v_1,\ldots,v_N)' \sim (0,\ I_N \otimes \Delta)$, and $\varepsilon = (\varepsilon_{11},\ldots,\varepsilon_{1T},\ldots,\varepsilon_{NT})' \sim (0,\ \begin{bmatrix}\sigma_1^2 & & \\ & \ddots & \\ & & \sigma_N^2\end{bmatrix} \otimes I_T)$. Equation (4.7) becomes

$$(4.16)\qquad \tilde{\beta}_i = (\Delta^{-1} + \frac{1}{\sigma_i^2} x_i'x_i)^{-1}(\frac{1}{\sigma_i^2} x_i'x_i\hat{\beta}_i + \Delta^{-1}\tilde{\Theta})$$

where $x_i = (x_{i1}',\ldots,x_{iT}')'$ and $\hat{\beta}_i = (x_i'x_i)^{-1}x_i'y_i$ is the least squares estimator (LSE) based only on boservations for the i^{th} individual. The BLU predictor $\tilde{\beta}_i$ is a weighted average of the LSE $\hat{\beta}_i$ and a common (estimated) mean, with the weights related to the least squares variance matrix.[3] The coveriance matrix of $\tilde{\beta}_i - \beta_i$ is

$$\begin{aligned} &E(\tilde{\beta}_i - \beta_i)(\tilde{\beta}_i - \beta_i)' \\ &= (I-\Delta x_i'(x_i\Delta x_i' + \sigma_i^2 I_T)^{-1}x_i)(\Delta + (\Sigma_{j=1}^n x_j'(x_j\Delta x_j' + \sigma_j^2 I_T)^{-1}x_j)^{-1}) \cdot \\ &\quad (I-x_i'(x_i\Delta x_i' + \sigma_i^2 I_T)^{-1}x_i\Delta) + \sigma_i^2\Delta x_i'(x_i\Delta x_i' + \sigma_i^2 I_T)^{-2}x_i\Delta \qquad i=1,\ldots,N \end{aligned}$$

[3]The second author is indebted to George Battese for pointing out that the BLU properties of $\tilde{\beta}_i$ follow from Goldberger's (1962) result and hence that a rather lengthy proof in an earlier paper (Griffiths, 1974) was unnecessary. On this point, see also Rosenberg (1973b).

Swamy (1970, 1971) gives $\hat{\beta}_i$ as the BLU predictor of β_i. This is true if the class of unbiased predictors $\{\beta_i^*\}$ is restricted to those for which $E(\beta_i^*|\beta_i) = \beta_i$. We have defined an unbiased predictor as one for which $E(\beta_i^* - \beta_i) = 0$, where the expectation is an unconditional one. That is, we envisage repeated sampling over both time and over cross sectional units. In this context it is not difficult to show that the covariance matrix of $(\tilde{\beta}_i - \beta_i)$ is "less" than that of $(\hat{\beta}_i - \beta_i)$.

Equation (4.12) becomes

$$(4.17) \qquad \tilde{\theta} = (\Sigma_{j=1}^{n} x_j'(x_j \Delta x_j' + \sigma_j^2 I_T)^{-1} x_j)^{-1} \Sigma_{j=1}^{n} x_j'(x_j \Delta x_j' + \sigma_j^2 I_T)^{-1} y_j$$

which, of course, is the GLS estimator derived in Swamy (1970). As shown in Swamy (1970), $\tilde{\theta}$ is a weighted average of the LSE's $\hat{\beta}_i$, i=1,...,N, with weights given by the precision matrices of the $\hat{\beta}_i$,

$$\tilde{\theta} = (\Sigma_{i=1}^{n}(\Delta + \sigma_i^2(x_i'x_i)^{-1}))^{-1} \; \Sigma_{j=1}^{n}(\Delta + \sigma_j^2(x_j'x_j)^{-1})^{-1}\hat{\beta}_j \; .$$

From (4.8), equivalently, $\tilde{\theta}$ is simply an arithmetical average of the $\tilde{\beta}_i$,

$$\tilde{\theta} = \frac{1}{N}\Sigma_{i=1}^{n}\tilde{\beta}_i \; .$$

Other models, such as those of Hildreth and Houck (1968) and Hsiao (1975), can also be framed in terms of (4.1) and (4.2) and the relevant estimators derived. For the Hildreth-Houck model the resulting BLUE of the random components is the same as that given by Griffiths (1972).

5. Bayesian Estimation with Known Covariances

If one assumes normal disturbances, known variances, a diffuse prior on the constant underlying parameters, and quadratic loss, so that the posterior parameter means will be the "Bayesian estimators", it is not difficult to show that these estimators will be identical to those obtained from the prior likelihood approach. For the general model in equations (2.5) and (2.6) we can write the posterior density for the unknown parameters as

$$(5.1) \qquad p(\beta_1,\beta_2, \Theta|y, \Sigma, \Omega) \propto \ell(y|\beta_1,\beta_2,\Sigma)p_1(\beta_1|\Theta, \Omega)p_2(\Theta, \beta_2)$$

where $\ell(\cdot)$ is the likelihood function of y, $p_1(\cdot)$ is the density function of β_1 and $p_2(\cdot)$ is the prior density function for the parameters (Θ, β_2). If $p_2(\Theta, \beta_2) \propto$ constant (i.e., a diffuse prior), the posterior $p(\beta_1, \beta_2, \Theta|y, \Sigma, \Omega)$ will be algebraically identical to the "likelihood" in equation (2.8). If in this function, one completes the squares on β_1, β_2 and Θ, it is evident that $p(\beta_1,\beta_2, \Theta|y, \Sigma, \Omega)$ is a multivariate normal distribution and hence that the marginals $p(\beta_1|y, \Sigma, \Omega)$, $p(\beta_2|y, \Sigma, \Omega)$ and $p(\Theta|y, \Sigma, \Omega)$ are also normal distributions. Further, their means will be equal to their respective modes and it is the modes which were the maximum likelihood estimators obtained above.

Thus, in a Bayesian context, Smith (1973) obtained the result for $\tilde{\beta}$ given in equation (4.7) and also derived the result for the Swamy random coefficient model.[4] See also Leamer (1978, p. 274) and Rosenberg (1973b). The corresponding estimator for the Hildreth-Houck random coefficient model was obtained by Griffiths et. al. (1979).

---

[4] Smith (1973) derived his results for the general model as in (4.1) and (4.2) with the restriction that X and A have full ranks. The full rank condition restricts the scope of applications and is unnecessary. For example, Hsiao's model (1975) does not satisfy such conditions.

6. Analysis with Unknown Variances

The above analysis assumed that variances are known. If variances are unknown, the results can be approximated with estimated variances. In the variance components models, consistent estimators for the variances σ_u^2, σ^2 and σ_v^2, in (3.1) or (3.20), have been given by Graybill (1961), Wallace and Hussain (1964) and others. In the random coefficient model in (4.15), consistent estimators of σ_i^2 and Δ can be found in Swamy (1971). Swamy (1971) has also proposed unbiased estimators (see Swamy (1971), p. 107). These consistent estimates can be substituted into the relevant expressions derived above. Under general conditions imposed on the exogenous variables in the specific models, the estimators for the fixed parameters (in (3.8) and (3.28) for the variance components models and (4.17) for the random coefficient model) will be consistent, asymptotically normal and efficient when both N and T tend to infinity. For the estimators (predictors) of the individual components it is obvious that as N and T tend to infinity the sample information, for a given unit, will dominate the additional weak stochastic information. It follows that the estimators of the random coefficients in (3.7), (3.26) and (3.27) for the variance components models, and (4.16) for the random coefficient model, will be derived from the samples only, namely, $\tilde{u}_i \approx \bar{y}_i - \bar{x}_i\tilde{\beta}$ in (3.7), $\tilde{\alpha}_i \approx \bar{y}_i - \bar{x}_i\tilde{\beta}$, $\tilde{\gamma}_t \approx \bar{y}_{\cdot t} - \bar{x}_{\cdot t}\tilde{\beta} - \tilde{\alpha}$ in (3.26), (3.27) and $\tilde{\beta}_i \approx \hat{\beta}_i$ in (4.16). Thus, it is easy to see that in the limit, $\operatorname*{plim}_{T\to\infty} \tilde{u}_i = u_i$ for variance component model (3.1), $\operatorname*{plim}_{T\to\infty} \tilde{\alpha}_i = \alpha_i$ and $\operatorname*{plim}_{N\to\infty} \tilde{\gamma}_t = 0$ for model (3.20), and $\operatorname*{plim}_{T\to\infty} \tilde{\beta}_i = \beta_i$ for the random coefficients model (4.15).

When all the disturbances are normally distributed, the prior likelihood approach provides an alternative method of estimating the covariances in the variance component and random coefficient models. For the variance components model in (3.20), the log likelihood function incorporating the prior likelihood is

$$(6.1) \qquad \ln L(\alpha,\beta,\alpha_i,\gamma_t,\sigma_u^2,\sigma_v^2,\sigma^2|y,X)$$

$$= \text{constant} - \frac{NT}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\Sigma_{i=1}^{N}\Sigma_{i=1}^{T}(y_{it}-x_{it}\beta-\alpha_i-\gamma_t)^2 - \frac{N}{2}\ln\sigma_u^2$$

$$- \frac{1}{2\sigma_u^2}\Sigma_{i=1}^{N}(\alpha_i-\alpha)^2 - \frac{T}{2}\ln\sigma_v^2 - \frac{1}{2\sigma_v^2}\Sigma_{t=1}^{T}\gamma_t^2 .$$

This suggests that estimates can be derived from the following system of equations,

$$(6.2) \qquad \tilde{\beta} = \Big(\frac{1}{NT}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}x_{it}'x_{it} - \frac{T\tilde{\sigma}_u^2}{T\tilde{\sigma}_u^2+\tilde{\sigma}^2}\frac{1}{N}\Sigma_{i=1}^{N}\bar{x}_i'\bar{x}_i - \frac{N\tilde{\sigma}_v^2}{N\tilde{\sigma}_v^2+\tilde{\sigma}^2} \cdot$$

$$\frac{1}{T}\Sigma_{t=1}^{T}\bar{x}_{\cdot t}'\bar{x}_{\cdot t}\Big)^{-1}\Big(\frac{1}{NT}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}x_{it}'y_{it} - \frac{T\tilde{\sigma}_u^2}{T\tilde{\sigma}_u^2+\tilde{\sigma}^2}\frac{1}{N}\Sigma_{i=1}^{N}\bar{x}_i'\bar{y}_i -$$

$$\frac{N\tilde{\sigma}_v^2}{N\tilde{\sigma}_v^2+\tilde{\sigma}^2}\frac{1}{T}\Sigma_{t=1}^{T}\bar{x}_{\cdot t}'\bar{y}_{\cdot t}\Big)$$

$$\tilde{\alpha}_i = \frac{1}{T\tilde{\sigma}_u^2+\tilde{\sigma}^2}\left(T\tilde{\sigma}_u^2(\bar{y}_i-\bar{x}_i\tilde{\beta}) + \tilde{\sigma}^2\bar{y}\right) \qquad i=1,\ldots,N$$

$$\tilde{\gamma}_t = \frac{N\tilde{\sigma}_v^2}{N\tilde{\sigma}_v^2+\tilde{\sigma}^2}(\bar{y}_{\cdot t} - \bar{x}_{\cdot t}\tilde{\beta} - \bar{y}) \qquad t=1,\ldots,T$$

and

$$(6.3)\qquad \tilde{\sigma}^2 = \frac{1}{NT}\,\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}(y_{it}-x_{it}\tilde{\beta} - \tilde{\alpha}_i - \tilde{\gamma}_t)^2$$

$$\tilde{\sigma}_u^2 = \frac{1}{N}\;\Sigma_{i=1}^{N}(\tilde{\alpha}_i - \bar{y})^2$$

$$\tilde{\sigma}_u^2 = \frac{1}{T}\;\Sigma_{i=1}^{T}\tilde{\gamma}_t^2$$

Equations in (6.2) and (6.3) form a block recursive nonlinear equations system. With estimated variances, the estimates of the coefficients are derived from (6.2) and with estimated coefficients, estimates of the variances can be derived from (6.3). The iterative solutions from (6.2) and (6.3) satisfy the first order conditions $\frac{\partial \ell nL}{\partial \Theta} = 0$, where $\Theta = (\alpha,\beta,\alpha_i,\gamma_t,\sigma_u^2,\sigma_v^2,\sigma^2)'$ from (6.1), are the likelihood roots, and hence are the maximum likelihood estimates. The consistency of these likelihood roots is guaranteed when the iterations begin with consistent estimates of the variances such as those provided in the literature. It should be noted that one should avoid using a search algorithm to find the mode (global maximum) of the likelihood function in (6.1) which is, in fact, unbounded from above. In this case, the consistent estimates correspond to a local maximum. The unboundedness occurs at the parameter values $\alpha_i = \alpha$ for all $i = 1,\ldots,N$ and $\sigma_u^2 = 0$, or $\gamma_t = 0$ for all t and $\sigma_v^2 = 0$.

For the random coefficients model in (4.15), the log likelihood function is

(6.4) $$\ell n\, L(\Theta, \beta_i, \sigma_i^2, \Delta | y, x)$$

$$= \text{constant} - \frac{T}{2}\Sigma_{i=1}^{N} \ell n \sigma_i^2 - \frac{1}{2}\Sigma_{i=1}^{N} \frac{1}{\sigma_i^2}(y_i - x_i\beta_i)'(y_i - x_i\beta_i) -$$

$$\frac{N}{2}\, \ell n|\Delta| - \frac{1}{2}\Sigma_{i=1}^{N}(\beta_i - \beta)'\Delta^{-1}(\beta_i - \beta) .$$

The recursive equations system is

(6.5) $$\tilde{\Theta} = (\Sigma_{i=1}^{n}(\tilde{\Delta} + \tilde{\sigma}_i^2(x_i'x_i)^{-1}))^{-1}\Sigma_{j=1}^{n}(\tilde{\Delta} + \tilde{\sigma}_j^2(x_j'x_j)^{-1})^{-1}(x_j'x_j)^{-1}x_j'y_j$$

$$\tilde{\beta}_i = \tilde{\Theta} + \tilde{\Delta}x_i'\ (x_i\tilde{\Delta}x_i' + \tilde{\sigma}_i^2 I_T)^{-1}(y_i - x_i\tilde{\Theta}) \qquad i=1,\ldots,N$$

and

(6.6) $$\tilde{\Delta} = \frac{1}{N}\,\Sigma_{i=1}^{N}(\tilde{\beta}_i - \tilde{\Theta})(\tilde{\beta}_i - \tilde{\Theta})'$$

$$\tilde{\sigma}_i^2 = \frac{1}{T}(y_i - x_i\tilde{\beta}_i)'(y_i - x_i\tilde{\beta}_i) \qquad i=1,\ldots,N.$$

The iterative solutions correspond to the likelihood roots. If consistent estimators for the variances such as

(6.7) $$\hat{\Delta} = \frac{1}{N}\,\Sigma_{j=1}^{N}(\hat{\beta}_j - \frac{1}{N}\,\Sigma_{i=1}^{N}\hat{\beta}_i)(\hat{\beta}_j - \frac{1}{N}\Sigma_{i=1}^{N}\hat{\beta}_i)',$$

$$\hat{\sigma}_i^2 = \frac{1}{T}\,(y_i - x_i\hat{\beta}_i)'(y_i - x_i\hat{\beta}_i) ,$$

where $\hat{\beta}_i = (x_i'x_i)^{-1}x_i'y_i$ is the least squares estimator, are used to provide initial estimates, the iterative solutions from (6.5) and (6.6) will be consistent. The ultimate estimator $\tilde{\Delta}$ is expected to be better than $\hat{\Delta}$ since it utilizes more information. Also, $\tilde{\Delta}$ in (6.6) is guaranteed nonnegative definite and hence avoids the possibility of negative variance estimates. Similar remarks can be applied to $(\tilde{\sigma}_u^2, \tilde{\sigma}_v^2)$ in (6.3) for the variance components models.

Now let us compare the Bayesian approach when the disturbances are unknown. Consider the general model in equations (2.5) and (2.6). In this case, one need to assign a joint prior distribution to Θ, β_2, Σ and Ω. The posterior density for the unknown parameters is

$$(6.8) \qquad p(\beta_1,\beta_2,\Theta,\Sigma,\Omega|y) \propto \ell(y|\beta_1,\beta_2,\Sigma)p_1(\beta_1|\Theta,\Omega)\ p_2(\Theta,\beta_2,\Sigma,\Omega)$$

where $p_2(\Theta, \beta_2,\Sigma,\Omega)$ is the joint prior distribution of Θ, β_2, Σ and Ω. Since the parameters β_1, β_2, Θ are the parameters of interest, the parameters Σ and Ω, as nuisance parameters, need to be integrated out of the posterior density. The mean (for quadratic loss) of the resulting marginal posterior $p(\beta_1,\beta_2,\Theta|y)$ gives the Bayesian estimates. However, in most of the models above and for any reasonable prior distribution p_2, integration with respect to Σ and Ω is analytically intractible. Thus the posterior means can only be obtained by numerical integration which, computationally, is limited to a small number of dimensions. See discussions in Lindley (1971), Lindley and Smith (1972) and Griffiths et. al. (1979). Also, it follows, that when the variances are unknown, the formal solutions from the prior likelihood approach and the Bayesian approach are different.

Instead of the full Bayesian analysis, Lindley (1971) and Lindley and Smith (1972) suggest approximations: the first approximation consists of using the mode of the posterior distribution in place of the mean and, secondly the use of the mode of the joint distribution rather than that of the $(\beta_1, \beta_2, \Theta)$ - margin. Thus, mathematically, the solutions are derived by finding the mode of the joint distribution p in (6.8). The difference between this solution and the prior likelihood solution will then depend on the choice of the prior distribution p_2. However, with complete ignorance, the choice of p_2 involves a technical difficulty. When p_2 is a diffuse prior, the posterior distribution for β_1, β_2, Θ, Σ and Ω is not proper (Lindley (1971)) and p in (6.8) is unbounded. To overcome this difficulty, Lindley and Smith suggest a proper prior distribution, namely, for the random coefficients model in (4.15), the conjugate Wishart distribution for Δ^{-1}, independent, conjugate inverse $-\chi^2$ distributions for the σ_i's and a diffuse prior for Θ. The resulting model estimates, as reported in Smith (1973), satisfy a recursive equation system similar to equations in (6.5) and (6.6) with (6.6) modified as

$$\Delta^* = \frac{1}{N-K-2+\rho}\{R + \Sigma_{i=1}^{N}(\beta_i^* - \Theta^*)(\beta_i^* - \Theta^*)'\}$$

$$\sigma_i^{2*} = \frac{1}{T + \nu_i + 2}\{\nu_i\lambda_i + (y_i - x_i\beta_i^*)'(y_i - x_i\beta_i^*)\} \qquad i=1,\ldots,N$$

where β_i^*, etc. are the model estimates, and $(\rho, R, \nu_i, \lambda_i)$ are parameters arising in the prior specification. In empirical applications, they suggest taking small positive values for λ_i and small entries for R to approximate vague priors. The approximate solutions will, thus be similar to the likelihood approach.

7. Concluding Remarks

Alternative regression models which make various assumptions about random components are currently very popular in the econometric literature. In this paper, for such models, we have demonstrated the usefulness and potential of the prior likelihood approach, particularly with respect to estimation of random components. A number of interesting questions remain. In particular, when the variances are unknown, the small sample properties of the various estimators are not available. Monte Carlo experiments may shed some light on whether or not the prior likelihood variance estimators are better, or lead to better estimates of the other coefficients, than some of the traditional variance estimators.

References

Balestra, P. and M. Nerlove (1966), "Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas", Econometrica, 34, pp. 585-612.

Chipman, J. S. (1964), "On Least Squares with Insufficient Observations", Journal of the American Statistical Association, 59, pp. 1078-1111.

Cooley, T. and E. C. Prescott (1973), "An Adaptive Regression Model", International Economic Review.

DeFinetti, B. (1964), "Foresight: Its Logical Laws, Its Subjective Sources", in Studies in Subjective Probability, H. E. Kyburg, Jr. and H. E. Smokler, eds., pp. 93-158, New York: Wiley.

Edwards, A. W. F. (1969), "Statistical Methods in Scientific Inference", Nature, Land 222, pp. 1233-1237.

Goldberger, A. S. (1962), "Best Linear Unbiased Prediction in the Generalized Linear Regression Model", Journal of the American Statistical Association, 57, pp. 369-375.

Graybill, F. A. (1961), An Introduction to Linear Statistical Models, McGraw-Hill Book Company, Inc. New York.

Griffiths, W. E. (1972), "Estimating Actual Response Coefficients in the Hildreth-Houck Random Coefficient Model", Journal of the American Statistical Association, 67, pp. 633-635.

Griffiths, W. E. (1974), "Combining Time Series-Cross Section Data: Alternative Models and Estimators", paper presented at the Fourth Conference of Economists, Canberra.

Griffiths, W. E., R. G. Dryman and S. Prakash (1979), "Bayesian Estimation of A Random Coefficient Model", Manuscript, University of New England, forthcoming in Journal of Econometrics.

Hildreth, C., and J. Houck (1968), "Some Estimators for a Linear Model with Random Coefficients", Journal of the American Statistical Association, 63, pp. 584-595.

Hsiao, Cheng (1975), "Some Estimation Methods for a Random Coefficient Model", Econometrica, 43, pp. 305-325.

Learner, E. E., Specification Searches, John Wiley and Sons, New York, 1978.

Lindley, D. V. (1971), "The Estimation of Many Parameters", in Foundations of Statistical Inference, V. P. Godambe and D. A. Sprott, eds., pp. 435-455. Toronto: Holt, Rinehart and Winston.

Lindley, D. V. and Smith, A.F.M. (1972), "Bayes Estimates for the Linear Model (with Discussion)", J. R. Statist. Soc. B, 34, pp. 1-41.

Maddala, G. S. (1971), "The Use of Variance Components Models in Pooling Cross Section and Time Series Data", Econometrica, 39, pp. 341-358.

Nelder, J. A. (1972), Discussion of a paper by D. V. Lindley and A.F.M. Smith on Bayes Estimates, J. R. Statist. Soc. B, 34, pp. 18-20.

Nerlove, M. (1971), "A Note on Error Components Models", Econometrica 39, pp. 383-396.

Rosenberg, B. (1973a), "The Analysis of a Cross-Section of Time Series by Stochastically Convergent Parameter Regression", Annals of Economic and Social Measurement, pp. 399-450.

Rosenberg, B. (1973b), "Linear Regression with Randomly Dispersed Parameters", Biometrika 60, pp. 65-72.

Smith, A.F.M. (1973), "A General Bayesian Linear Model", J. R. Statist. Soc. B, 35, pp. 67-75.

Swamy, P.A.V.B. (1970), "Efficient Inference in a Random Coefficient Regression Model", Econometrica, 33, pp. 311-323.

Swamy, P.A.V.B. (1971), Statistical Inference in Random Coefficient Regression Models, Springer-Verlag, New York.

Theil, H., and A. S. Goldberger (1960), "On Pure and Mixed Statistical Estimation in Economics", International Economics Review, 2, pp. 65-78.

Theil, H. (1963), "On the Use of Incomplete Prior Information in Regression Analysis", Journal of the American Statistical Association, 58, pp. 401-414.

Theil, H. (1971), Principles of Econometrics, New York, John Wiley and Sons.

Wallace, T. D., and Ashiq Hussain (1969), "The Use of Error Components Models in Combining Cross-Section with Time Series Data", Econometrica, 37, pp. 55-68.