

Random Coefficient Panel Data Models*

Cheng Hsiao

University of Southern California

M. Hashem Pesaran

University of Cambridge & USC

June 2004

Abstract

This paper provides a review of linear panel data models with slope heterogeneity, introduces various types of random coefficients models and suggest a common framework for dealing with them. It considers the fundamental issues of statistical inference of a random coefficients formulation using both the sampling and Bayesian approaches. The paper also provides a review of heterogeneous dynamic panels, testing for homogeneity under weak exogeneity, simultaneous equation random coefficient models, and the more recent developments in the area of cross-sectional dependence in panel data models.

Keywords: Random coefficient models, Dynamic heterogeneous panels, Classical and Bayesian approaches, Tests of slope heterogeneity, Cross section dependence.

JEL-Classification: C12, C13, C33.

*We are grateful to G. Bressons and A. Pirotte for their careful reading of an early version and for pointing out many typos. We would also like to thank J. Breitung and Ron Smith for helpful comments.

1 Introduction

Consider a linear regression model of the form

$$y = \boldsymbol{\beta}' \mathbf{x} + u, \quad (1)$$

where y is the dependent variable and \mathbf{x} is a $K \times 1$ vector of explanatory variables. The variable u denotes the effects of all other variables that affect the outcome of y but are not explicitly included as independent variables. The standard assumption is that u behaves like a random variable and is uncorrelated with \mathbf{x} . However, the emphasis of panel data is often on the individual outcomes. In explaining human behavior, the list of relevant factors may be extended ad infinitum. The effect of these factors that have not been explicitly allowed for may be individual specific and time varying. In fact, one of the crucial issues in panel data analysis is how the differences in behavior across individuals and/or through time that are not captured by \mathbf{x} should be modeled.

The variable intercept and/or error components models attribute the heterogeneity across individuals and/or through time to the effects of omitted variables that are individual time-invariant, like sex, ability and social economic background variables that stay constant for a given individual but vary across individuals, and/or period individual-invariant, like prices, interest rates and wide spread optimism or pessimism that are the same for all cross-sectional units at a given point in time but vary through time. It does not allow the interaction of the individual specific and/or time varying differences with the included explanatory variables, \mathbf{x} . A more general formulation would be to let the variable y of the individual i at time t be denoted as

$$\begin{aligned} y_{it} &= \boldsymbol{\beta}'_{it} \mathbf{x}_{it} + u_{it}, \\ &= \beta_{1it} x_{1it} + \dots + \beta_{kit} x_{kit} + u_{it}, \end{aligned} \quad (2)$$

$i = 1, \dots, N$, and $t = 1, \dots, T$. Expression (2) corresponds to the most general specification of the panel linear data regression problem. It simply states that each individual has their own coefficients that are specific to each time period. However, as pointed out by Balestra (1991) this general formulation is, at most, descriptive. It lacks any explanatory power and it is useless for prediction. Furthermore, it is not estimable, the number of parameters to be estimated exceeds

the number of observations. For a model to become interesting and to acquire explanatory and predictive power, it is essential that some structure is imposed on its parameters.

One way to reduce the number of parameters in (2) is to adopt an analysis of variance framework by letting

$$\beta_{kit} = \beta_k + \alpha_{ki} + \lambda_{kt}, \quad \sum_{i=1}^N \alpha_{ki} = 0, \quad \text{and} \quad \sum_{t=1}^T \lambda_{kt} = 0, \quad k = 1, \dots, K. \quad (3)$$

This specification treats individual differences as fixed and is computationally simple. The drawback is that it is not parsimonious, and hence reliable estimates of α_{ki} and λ_{kt} are difficult to obtain. Moreover, it is difficult to draw inference about the population if differences across individuals and/or over time are fixed and different.

An alternative to the fixed coefficient (or effects) specification of (3) is to let α_{ki} and λ_{kt} be random variables and introduce proper stochastic specifications. This is commonly called the “random coefficients” model. The random coefficient specification reduces the number of parameters to be estimated substantially, while still allowing the coefficients to differ from unit to unit and/or from time to time.

In section 2 we introduce various types of random coefficients models and suggest a common framework for them. In sections 3 and 4 we consider the fundamental issues of statistical inference of a random coefficients formulation using the sampling approach. In section 5 we consider a Bayesian approach. Section 6 considers the generalization to a dynamic framework. Issues of testing for homogeneity under weak exogeneity are discussed in section 7. Discussions on random coefficients, simultaneous equation systems and cross-sectional dependence are in sections 8 and 9, respectively. Conclusions are in section 10.

2 The Models

Let there be observations for N cross-sectional units over T time periods. Suppose the variable y for the i th unit at time t is specified as a linear function of K strictly

exogenous variables, $x_{kit}, k = 1, 2, \dots, K$, in the form¹

$$\begin{aligned} y_{it} &= \sum_{k=1}^K \beta_{kit} x_{kit} + u_{it}, \\ &= \boldsymbol{\beta}'_{it} \mathbf{x}_{it} + u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \end{aligned} \quad (4)$$

where u_{it} denotes the random error term, \mathbf{x}_{it} is a $K \times 1$ vector of exogenous variables and $\boldsymbol{\beta}_{it}$ is the $K \times 1$ vector of coefficients. The random coefficients approach assumes that the coefficients $\boldsymbol{\beta}_{it}$ are draws from probability distributions with a fixed number of parameters that do not vary with N and/or T . Depending on the type of assumption about the parameter variation, we can further classify the models into one of two categories: stationary and non-stationary random-coefficients models.

The stationary random-coefficients models regard the coefficients as having constant means and variance-covariances. Namely, the $K \times 1$ vector $\boldsymbol{\beta}_{it}$ is specified as

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta} + \boldsymbol{\xi}_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (5)$$

where $\boldsymbol{\beta}$ is a $K \times 1$ vector of constants, and $\boldsymbol{\xi}_{it}$ is a $K \times 1$ vector of stationary random variables with zero means and constant variance-covariances. For instance, in the Swamy (1970) type random coefficient models,

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta} + \boldsymbol{\alpha}_i, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (6)$$

and

$$\begin{aligned} E(\boldsymbol{\alpha}_i) &= \mathbf{0}, \quad E(\boldsymbol{\alpha}_i \mathbf{x}'_{it}) = \mathbf{0}, \\ E(\boldsymbol{\alpha}_i \boldsymbol{\alpha}'_j) &= \begin{cases} \boldsymbol{\Delta}, & \text{if } i = j, \\ \mathbf{0}, & \text{if } i \neq j, \end{cases} \end{aligned} \quad (7)$$

Hsiao (1974, 1975) considers the following type of model

$$\begin{aligned} \boldsymbol{\beta}_{it} &= \boldsymbol{\beta} + \boldsymbol{\xi}_{it} \\ &= \boldsymbol{\beta} + \boldsymbol{\alpha}_i + \boldsymbol{\lambda}_t, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \end{aligned} \quad (8)$$

¹The case where one or more of the regressors are weakly exogenous is considered in Section 6.

and assumes

$$\begin{aligned}
E(\boldsymbol{\alpha}_i) &= E(\boldsymbol{\lambda}_t) = \mathbf{0}, E(\boldsymbol{\alpha}_i \boldsymbol{\lambda}'_t) = \mathbf{0}, \\
E(\boldsymbol{\alpha}_i \mathbf{x}'_{it}) &= \mathbf{0}, E(\boldsymbol{\lambda}_t \mathbf{x}'_{it}) = \mathbf{0}, \\
E(\boldsymbol{\alpha}_i \boldsymbol{\alpha}'_j) &= \begin{cases} \boldsymbol{\Delta}, & \text{if } i = j, \\ \mathbf{0}, & \text{if } i \neq j, \end{cases} \\
E(\boldsymbol{\lambda}_i \boldsymbol{\lambda}'_j) &= \begin{cases} \boldsymbol{\Lambda}, & \text{if } i = j, \\ \mathbf{0}, & \text{if } i \neq j, \end{cases}
\end{aligned} \tag{9}$$

Alternatively, a time varying parameter model may be treated as realizations of a stationary stochastic process, thus $\boldsymbol{\beta}_{it}$ can be written in the form,

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_t = \mathbf{H}\boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \tag{10}$$

where all eigenvalues of \mathbf{H} lie inside the unit circle, and $\boldsymbol{\eta}_t$ is a stationary random variable with mean $\boldsymbol{\mu}$. Then the Hildreth-Houck (1968) type model is obtained by letting $\mathbf{H} = \mathbf{0}$ and $\boldsymbol{\eta}_t$ be i.i.d.; for the Pagan (1980) model, $\mathbf{H} = \mathbf{0}$ and

$$\boldsymbol{\eta}_t - \boldsymbol{\mu} = \boldsymbol{\eta}_t - \bar{\boldsymbol{\beta}} = a(L)\boldsymbol{\epsilon}_t, \tag{11}$$

where $\bar{\boldsymbol{\beta}}$ is the mean of $\boldsymbol{\beta}_t$ and $a(L)$ is the ratio of polynomials of orders p and q in the lag operator $L(L\boldsymbol{\epsilon}_t = \boldsymbol{\epsilon}_{t-1})$ and $\boldsymbol{\epsilon}_t$ is independent normal. The Rosenberg (1972, 1973) return-to-normality model assumes the absolute value of the characteristic roots of \mathbf{H} be less than 1 with $\boldsymbol{\eta}_t$ independently normally distributed with mean $\boldsymbol{\mu} = (\mathbf{I}_K - \mathbf{H})\bar{\boldsymbol{\beta}}$.

The nonstationary random coefficients models do not regard the coefficient vector as having constant mean or variances. Changes in coefficients from one observation to the next can be the result of the realization of a nonstationary stochastic process or can be a function of exogenous variables. When the coefficients are realizations of a nonstationary stochastic process, we may again use (10) to represent such a process. For instance, the Cooley and Prescott (1976) model can be obtained by letting $\mathbf{H} = \mathbf{I}_K$ and $\boldsymbol{\mu} = \mathbf{0}$. When the coefficients $\boldsymbol{\beta}_{it}$ are functions of individual characteristics or time variables (e.g. Amemiya (1978), Boskin and Lau (1990), Hendricks, Koenker, and Poirier (1979), Singh et al. (1976), Swamy and Tinsley (1977), Wachter (1970)), we can let

$$\boldsymbol{\beta}_{it} = \boldsymbol{\Gamma}\mathbf{q}_{it} + \boldsymbol{\eta}_{it}. \tag{12}$$

While the detailed formulation and estimation of the random coefficients model depends on the specific assumptions about the parameter variation, many types of the random coefficients models can be conveniently represented using a mixed fixed and random coefficients framework of the form (e.g. Hsiao (1990), Hsiao, Appelbe and Dineen (1992))

$$\mathbf{y}_{it} = \mathbf{z}'_{it}\bar{\boldsymbol{\gamma}} + \mathbf{w}'_{it}\boldsymbol{\alpha}_{it} + u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (13)$$

where \mathbf{z}_{it} and \mathbf{w}_{it} are vectors of exogenous variables with dimensions ℓ and p respectively, $\bar{\boldsymbol{\gamma}}$ is an $\ell \times 1$ vector of constants, $\boldsymbol{\alpha}_{it}$ is a $p \times 1$ vector of random variables, and u_{it} is the error term. For instance, the Swamy type model ((6) and (7)) can be obtained from (13) by letting $\mathbf{z}_{it} = \mathbf{w}_{it} = \mathbf{x}_{it}$, $\bar{\boldsymbol{\gamma}} = \bar{\boldsymbol{\beta}}$, and $\boldsymbol{\alpha}_{it} = \boldsymbol{\alpha}_i$; the Hsiao type model (8) and (9) is obtained by letting $\mathbf{z}_{it} = \mathbf{w}_{it} = \mathbf{x}_{it}$, $\bar{\boldsymbol{\gamma}} = \bar{\boldsymbol{\beta}}$, and $\boldsymbol{\alpha}_{it} = \boldsymbol{\alpha}_i + \boldsymbol{\lambda}_t$; the stochastic time varying parameter model (10) is obtained by letting $\mathbf{z}_{it} = \mathbf{x}_{it}$, $\mathbf{w}'_{it} = \mathbf{x}'_{it}[\mathbf{H}, \mathbf{I}_K]$, $\bar{\boldsymbol{\gamma}} = \boldsymbol{\mu}$, and $\boldsymbol{\alpha}'_{it} = \boldsymbol{\lambda}'_t = [\boldsymbol{\beta}'_{t-1}, (\boldsymbol{\eta}_t - \boldsymbol{\mu})']$; and the model where $\boldsymbol{\beta}_{it}$ is a function of other variables (12) is obtained by letting $\mathbf{z}'_{it} = \mathbf{x}'_{it} \otimes \mathbf{q}'_{it}$, $\bar{\boldsymbol{\gamma}}' = \text{vec}(\Gamma)$, $\mathbf{w}_{it} = \mathbf{x}_{it}$, $\boldsymbol{\alpha}_{it} = \boldsymbol{\eta}_{it}$, etc.

For ease of illustrating the fundamental issues involved in estimating a random coefficients model we shall make the simplifying assumption that $\boldsymbol{\alpha}_{it} = \boldsymbol{\alpha}_i$ and $\boldsymbol{\alpha}_i$ are independently normally distributed over i with mean $\mathbf{0}$ and covariance $\boldsymbol{\Delta}$, denoted by $\boldsymbol{\alpha}_i \sim N(\mathbf{0}, \boldsymbol{\Delta})$. In other words, there are only individual-specific effects, $\boldsymbol{\alpha}_i$, and these individual - specific effects stay constant over time. Under this simplified assumption, model (13) can be written in the stacked form

$$\mathbf{y} = \bar{\mathbf{Z}}\bar{\boldsymbol{\gamma}} + \mathbf{W}\boldsymbol{\alpha} + \mathbf{u}, \quad (14)$$

where

$$\begin{aligned}
\mathbf{y}_{NT \times 1} &= \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{pmatrix}, \quad \mathbf{y}_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}, \quad \mathbf{u}_{NT \times 1} = \begin{pmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_N \end{pmatrix}, \quad \mathbf{u}_i = \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix}, \\
\bar{\mathbf{Z}}_{NT \times \ell} &= \begin{pmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_N \end{pmatrix}, \quad \mathbf{Z}_i = \begin{pmatrix} \mathbf{z}'_{i1} \\ \vdots \\ \mathbf{z}'_{iT} \end{pmatrix}, \\
\mathbf{W}_{NT \times Np} &= \begin{pmatrix} \mathbf{W}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 & \cdots & \mathbf{0} \\ \vdots & & \ddots & \\ \mathbf{0} & & & \mathbf{W}_N \end{pmatrix}, \quad \mathbf{W}_i = \begin{pmatrix} \mathbf{w}'_{i1} \\ \vdots \\ \mathbf{w}'_{iT} \end{pmatrix}, \quad \text{and } \boldsymbol{\alpha}_{Np \times 1} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_N \end{pmatrix}.
\end{aligned} \tag{15}$$

We further assume that $\boldsymbol{\alpha}$ and \mathbf{u} are mutually independent with

$$E(\mathbf{u}) = \mathbf{0}, \text{ and } E(\mathbf{u}\mathbf{u}') = \mathbf{C}. \tag{16}$$

3 Sampling Approach

Let

$$\mathbf{v} = \mathbf{W}\boldsymbol{\alpha} + \mathbf{u}, \tag{17}$$

then $E(\mathbf{v}) = \mathbf{0}$ and

$$E(\mathbf{v}\mathbf{v}') = \mathbf{W}(\mathbf{I}_N \otimes \boldsymbol{\Delta})\mathbf{W}' + \mathbf{C} = \boldsymbol{\Omega}. \tag{18}$$

Model (14) can be viewed as a linear regression model of the form

$$\mathbf{y} = \bar{\mathbf{Z}}\bar{\boldsymbol{\gamma}} + \mathbf{v}, \tag{19}$$

where the composite error term, \mathbf{v} , has a nonspherical covariance matrix. From a sampling point of view, the interest for model (19) will lie in (a) estimating the mean coefficient vector $\bar{\boldsymbol{\gamma}}$, (b) estimating the covariance matrix of \mathbf{v} , $\boldsymbol{\Omega}$, and (c) predicting y_{it} .

If $\mathbf{\Delta}$ and \mathbf{C} are known, the best linear unbiased estimator of $\bar{\boldsymbol{\gamma}}$ is the generalized least squares (GLS) estimator

$$\hat{\boldsymbol{\gamma}} = (\bar{\mathbf{Z}}' \boldsymbol{\Omega}^{-1} \bar{\mathbf{Z}})^{-1} (\bar{\mathbf{Z}}' \boldsymbol{\Omega}^{-1} \mathbf{y}), \quad (20)$$

with covariance matrix

$$\mathbf{D} = \text{Cov}(\hat{\boldsymbol{\gamma}}) = (\bar{\mathbf{Z}}' \boldsymbol{\Omega}^{-1} \bar{\mathbf{Z}})^{-1}. \quad (21)$$

If $\mathbf{\Delta}$ and \mathbf{C} are unknown, we can apply a two step GLS estimator. In the first step we estimate $\mathbf{\Delta}$ and \mathbf{C} . In the second step we estimate $\bar{\boldsymbol{\gamma}}$ by substituting the estimated $\mathbf{\Delta}$ and \mathbf{C} into (20) and treating them as if they were known. Provided $\mathbf{\Delta}$ and \mathbf{C} can be consistently estimated, the two step GLS will have the same asymptotic efficiency as the GLS estimator.

Similarly, we can obtain the best linear unbiased predictor of y_{if} using the formula

$$\begin{aligned} \hat{y}_{if} &= \mathbf{z}'_{if} \bar{\boldsymbol{\gamma}} + E(v_{if} | \mathbf{v}), \\ &= \mathbf{z}'_{if} \bar{\boldsymbol{\gamma}} + \text{Cov}(v_{if}, \mathbf{v})' \text{Var}(\mathbf{v})^{-1} \mathbf{v}. \end{aligned} \quad (22)$$

Because $\bar{\boldsymbol{\gamma}}$ and \mathbf{v} are unknown, their estimated values, $\hat{\boldsymbol{\gamma}}$ and $\hat{\mathbf{v}} = \mathbf{y} - \bar{\mathbf{Z}} \hat{\boldsymbol{\gamma}}$ are substituted into (22) in practice.

Equations (20) - (22) provide a general principle for efficient inference of a random coefficients model. To illustrate relations to a specific type of random coefficients model, we consider a Swamy type model (4), (6) and (7), assuming that the regressors \mathbf{z}_{it} , are strictly exogenous.

Under the assumptions of Swamy (1971), we have

$$\bar{\mathbf{Z}} = \mathbf{X} \mathbf{A}, \mathbf{W} = \mathbf{X}, \bar{\boldsymbol{\gamma}} = \bar{\boldsymbol{\beta}}, \boldsymbol{\beta}_i = \bar{\boldsymbol{\beta}} + \boldsymbol{\alpha}_i, \quad (23)$$

where

$$\begin{aligned} \mathbf{A}_{NT \times K} &= (\mathbf{I}_K, \mathbf{I}_K, \dots, \mathbf{I}_K)' \\ \mathbf{X}_{NT \times K} &= \begin{pmatrix} \mathbf{X}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{X}_N \end{pmatrix}, \mathbf{X}_i = \begin{pmatrix} \mathbf{x}'_{i1} \\ \vdots \\ \mathbf{x}'_{iT} \end{pmatrix}. \end{aligned} \quad (24)$$

For simplicity, we also assume that u_{it} is *independently* distributed across i and over t with

$$E(u_{it}^2) = \sigma_i^2. \quad (25)$$

Then $\mathbf{\Omega}$ is block diagonal, with the i^{th} diagonal block equal to

$$\mathbf{\Omega}_i = \mathbf{X}_i \mathbf{\Delta} \mathbf{X}_i' + \sigma_i^2 \mathbf{I}_T. \quad (26)$$

Substituting (23) - (26) into (20), the best linear unbiased estimator of the mean coefficient vector $\overline{\boldsymbol{\beta}}$ is

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{GLS} &= (\mathbf{A}' \mathbf{X}' \mathbf{\Omega}^{-1} \mathbf{X} \mathbf{A})^{-1} \mathbf{A}' \mathbf{X} \mathbf{\Omega}^{-1} \mathbf{y}, \\ &= \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{\Omega}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{\Omega}_i^{-1} \mathbf{y}_i \right), \\ &= \sum_{i=1}^N \mathbf{R}_i \widehat{\boldsymbol{\beta}}_i, \end{aligned} \quad (27)$$

where

$$\mathbf{R}_i = \left[\sum_{i=1}^N \left(\mathbf{\Delta} + \mathbf{\Sigma}_{\widehat{\boldsymbol{\beta}}_i} \right)^{-1} \right]^{-1} \left(\mathbf{\Delta} + \mathbf{\Sigma}_{\widehat{\boldsymbol{\beta}}_i} \right)^{-1}, \quad (28)$$

and

$$\widehat{\boldsymbol{\beta}}_i = (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{y}_i, \quad \mathbf{\Sigma}_{\widehat{\boldsymbol{\beta}}_i} = V(\widehat{\boldsymbol{\beta}}_i) = \sigma_i^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1}. \quad (29)$$

The last expression of (27) is obtained by repeatedly utilising the identity relation,

$$(\mathbf{E} + \mathbf{BFB}')^{-1} = \mathbf{E}^{-1} - \mathbf{E}^{-1} \mathbf{B} (\mathbf{B}' \mathbf{E}^{-1} \mathbf{B} + \mathbf{F}^{-1})^{-1} \mathbf{B}' \mathbf{E}^{-1}. \quad (30)$$

It shows that the GLS estimator is a matrix weighted average of the least squares estimator for each cross-sectional unit (29), with the weights inversely proportional to their covariance matrices. It also shows that the GLS estimator requires only a matrix inversion of order K , and so it is not much more complicated to compute than the sample least squares estimator.

The covariance matrix of the GLS estimator is

$$\text{Cov}(\widehat{\boldsymbol{\beta}}_{GLS}) = (\mathbf{A}' \mathbf{X}' \mathbf{\Omega}^{-1} \mathbf{X} \mathbf{A})^{-1} = \left[\sum_{i=1}^N \mathbf{X}_i' \mathbf{\Omega}_i^{-1} \mathbf{X}_i \right]^{-1} = \left[\sum_{i=1}^N \left(\mathbf{\Delta} + \mathbf{\Sigma}_{\widehat{\boldsymbol{\beta}}_i} \right)^{-1} \right]^{-1}. \quad (31)$$

If both errors and $\boldsymbol{\alpha}_i$ are normally distributed, the GLS estimator of $\bar{\boldsymbol{\beta}}$ is the maximum likelihood estimator (MLE) of $\bar{\boldsymbol{\beta}}$ conditional on $\boldsymbol{\Delta}$ and σ_i^2 . Without knowledge of $\boldsymbol{\Delta}$ and σ_i^2 , we can estimate $\bar{\boldsymbol{\beta}}$, $\boldsymbol{\Delta}$ and σ_i^2 , $i = 1, \dots, N$ simultaneously by the maximum likelihood method. However, computationally it can be tedious. A natural alternative is to first estimate $\boldsymbol{\Omega}_i$ then substitute the estimated $\boldsymbol{\Omega}_i$ into (27).

Swamy proposes using the least squares estimator of $\boldsymbol{\beta}_i$, $\hat{\boldsymbol{\beta}}_i = (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{y}_i$ and residuals $\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i$ to obtain unbiased estimators of σ_i^2 , $i = 1, \dots, N$, and $\boldsymbol{\Delta}$. Noting that

$$\hat{\mathbf{u}}_i = [\mathbf{I}_T - \mathbf{X}_i (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i] \mathbf{u}_i, \quad (32)$$

and

$$\hat{\boldsymbol{\beta}}_i = \boldsymbol{\beta}_i + (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{u}_i, \quad (33)$$

we obtain the unbiased estimators of σ_i^2 and $\boldsymbol{\Delta}$ as:

$$\begin{aligned} \hat{\sigma}_i^2 &= \frac{\hat{\mathbf{u}}'_i \hat{\mathbf{u}}_i}{T - K}, \\ &= \frac{1}{T - K} \mathbf{y}'_i [\mathbf{I}_T - \mathbf{X}_i (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i] \mathbf{y}_i, \end{aligned} \quad (34)$$

$$\begin{aligned} \hat{\boldsymbol{\Delta}} &= \frac{1}{N - 1} \sum_{i=1}^N \left(\hat{\boldsymbol{\beta}}_i - N^{-1} \sum_{j=1}^N \hat{\boldsymbol{\beta}}_j \right) \\ &\quad \left(\hat{\boldsymbol{\beta}}_i - N^{-1} \sum_{j=1}^N \hat{\boldsymbol{\beta}}_j \right)' - \frac{1}{N} \sum_{i=1}^N \hat{\sigma}_i^2 (\mathbf{X}'_i \mathbf{X}_i)^{-1}. \end{aligned} \quad (35)$$

Just as in the error-components model, the estimator (35) is not necessarily nonnegative definite. In this situation, Swamy [also see Judge et al. (1985)] has suggested replacing (35) by

$$\hat{\boldsymbol{\Delta}}^* = \frac{1}{N - 1} \sum_{i=1}^N \left(\hat{\boldsymbol{\beta}}_i - N^{-1} \sum_{j=1}^N \hat{\boldsymbol{\beta}}_j \right) \left(\hat{\boldsymbol{\beta}}_i - N^{-1} \sum_{j=1}^N \hat{\boldsymbol{\beta}}_j \right)'. \quad (36)$$

This estimator, although biased, is nonnegative definite and consistent when T tends to infinity.

4 Mean Group Estimation

A consistent estimator of $\bar{\beta}$ can also be obtained under more general assumptions concerning β_i and the regressors. One such possible estimator is the Mean Group (MG) estimator proposed by Pesaran and Smith (1995) for estimation of dynamic random coefficient models. The MG estimator is defined as the simple average of the OLS estimators, $\hat{\beta}_i$:

$$\hat{\beta}_{MG} = N^{-1} \sum_{i=1}^N \hat{\beta}_i. \quad (37)$$

When the regressors are strictly exogeneous and the errors, u_{it} are independently distributed, an unbiased estimator of the covariance matrix of $\hat{\beta}_{MG}$ can be computed as

$$\widehat{Cov}(\hat{\beta}_{MG}) = N^{-1} \hat{\Delta}^*,$$

where $\hat{\Delta}^*$ is given by (36). For a proof first note that under the random coefficient model we have

$$\hat{\beta}_i = \beta + \alpha_i + \xi_i,$$

where

$$\xi_i = (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{u}_i,$$

and

$$\hat{\beta}_{MG} = \beta + \bar{\alpha} + \bar{\xi},$$

where $\bar{\alpha} = \frac{1}{N} \sum_{i=1}^N \alpha_i$ and $\bar{\xi} = \frac{1}{N} \sum_{i=1}^N \xi_i$. Therefore

$$\hat{\beta}_i - \hat{\beta}_{MG} = (\alpha_i - \bar{\alpha}) + (\xi_i - \bar{\xi}),$$

$$\begin{aligned} (\hat{\beta}_i - \hat{\beta}_{MG}) (\hat{\beta}_i - \hat{\beta}_{MG})' &= (\alpha_i - \bar{\alpha}) (\alpha_i - \bar{\alpha})' + (\xi_i - \bar{\xi}) (\xi_i - \bar{\xi})' \\ &\quad + (\alpha_i - \bar{\alpha}) (\xi_i - \bar{\xi})' + (\xi_i - \bar{\xi}) (\alpha_i - \bar{\alpha})', \end{aligned}$$

and

$$\sum_{i=1}^N E \left[(\hat{\beta}_i - \hat{\beta}_{MG}) (\hat{\beta}_i - \hat{\beta}_{MG})' \right] = (N-1) \mathbf{\Delta} + \left(1 - \frac{1}{N}\right) \sum_{i=1}^N \sigma_i^2 (\mathbf{X}'_i \mathbf{X}_i)^{-1}.$$

But

$$\begin{aligned} Cov\left(\widehat{\boldsymbol{\beta}}_{MG}\right) &= Cov\left(\overline{\boldsymbol{\alpha}}\right) + Cov\left(\overline{\boldsymbol{\xi}}\right), \\ &= \frac{1}{N}\boldsymbol{\Delta} + \frac{1}{N^2}\sum_{i=1}^N\sigma_i^2 E\left[\left(\mathbf{X}'_i\mathbf{X}_i\right)^{-1}\right]. \end{aligned}$$

Using the above results it is now easily seen that

$$E\left[\widehat{Cov}\left(\widehat{\boldsymbol{\beta}}_{MG}\right)\right] = Cov\left(\widehat{\boldsymbol{\beta}}_{MG}\right),$$

as required.

Finally, it is worth noting that the MG and the Swamy estimators are in fact algebraically equivalent for T sufficiently large, namely

$$\lim_{T \rightarrow \infty} \left(\widehat{\boldsymbol{\beta}}_{GLS} - \widehat{\boldsymbol{\beta}}_{MG}\right) = \mathbf{0}.$$

5 Bayesian Approach

One can also derive the solutions for the model (14) from a Bayesian point of view. The Bayes approach assumes that all quantities, including the parameters, are random variables. Therefore, as part of the model, *prior* probability distributions are introduced for the parameters. The prior distribution is supposed to express a state of knowledge or ignorance about the parameters before the data are obtained. The model (14) with the assumption that $\overline{\boldsymbol{\gamma}}$ is fixed and $\boldsymbol{\alpha}_i$ is random, can be viewed as the state of knowledge about the parameters $\overline{\boldsymbol{\gamma}}$ and $\boldsymbol{\alpha}$ before the data are obtained: The prior distributions of $\overline{\boldsymbol{\gamma}}$ and $\boldsymbol{\alpha}$ are independent. There is no information on $\overline{\boldsymbol{\gamma}}$ but there is information on $\boldsymbol{\alpha}_i$, which is normally distributed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Delta}$. This prior information is combined with the model (14) and data, \mathbf{y} and \mathbf{z} , to revise the probability distribution of $\overline{\boldsymbol{\gamma}}$ and $\boldsymbol{\alpha}$, which is called the posterior distribution. From this distribution inferences are made.

Formally, we assume that

A1. The prior distributions of $\overline{\boldsymbol{\gamma}}$ and $\boldsymbol{\alpha}$ are independent, that is,

$$p(\overline{\boldsymbol{\gamma}}, \boldsymbol{\alpha}) = p(\overline{\boldsymbol{\gamma}}) \cdot p(\boldsymbol{\alpha}). \quad (38)$$

A2. There is no information about $\bar{\gamma}$,

$$p(\bar{\gamma}) \propto \text{constant}. \quad (39)$$

A3. There is prior information about α ,

$$\alpha \sim N(\mathbf{0}, \mathbf{I}_N \otimes \Delta). \quad (40)$$

Theorem 1. Suppose that, given $\bar{\gamma}$ and α ,

$$\mathbf{y} \sim N(\bar{\mathbf{Z}}\bar{\gamma} + \mathbf{W}\alpha, \mathbf{C}). \quad (41)$$

Under A1-A3,

(a) the marginal distribution of \mathbf{y} given $\bar{\gamma}$ is

$$\mathbf{y} \sim N(\bar{\mathbf{Z}}\bar{\gamma}, \mathbf{C} + \mathbf{W}(\mathbf{I}_N \otimes \Delta)\mathbf{W}'), \quad (42)$$

(b) the distribution of $\bar{\gamma}$ given \mathbf{y} is $N(\hat{\bar{\gamma}}, \mathbf{D})$, where $\hat{\bar{\gamma}}$ and \mathbf{D} are given by (20) and (21), respectively.

(c) the distribution of α given \mathbf{y} is $N(\hat{\alpha}, \tilde{\mathbf{D}})$, where

$$\begin{aligned} \hat{\alpha} = & \{\mathbf{W}'[\mathbf{C}^{-1} - \mathbf{C}^{-1}\bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\mathbf{C}^{-1}\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{C}^{-1}]\mathbf{W} + (\mathbf{I}_N \otimes \Delta^{-1})\}^{-1} \\ & \cdot \{\mathbf{W}'[\mathbf{C}^{-1} - \mathbf{C}^{-1}\bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\mathbf{C}^{-1}\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{C}^{-1}]\mathbf{y}\}, \end{aligned} \quad (43)$$

and

$$\tilde{\mathbf{D}} = \{\mathbf{W}'[\mathbf{C}^{-1} - \mathbf{C}^{-1}\bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\mathbf{C}^{-1}\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{C}^{-1}]\mathbf{W} + (\mathbf{I}_N \otimes \Delta^{-1})\}^{-1}. \quad (44)$$

See Appendix A for a proof.

Recall that

$$\beta = \mathbf{A}\bar{\beta} + \alpha, \quad (45)$$

and therefore the Bayes estimator of β can be obtained by substituting the Bayes estimators of $\bar{\beta}$ and α (27) and (43) into (45), namely:

$$\begin{aligned} \hat{\beta}^* &= \mathbf{A}\hat{\bar{\beta}}_{GLS} + \hat{\alpha} \\ &= \left(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X} + \tilde{\Delta}^{-1}\right)^{-1} \left(\mathbf{X}'\mathbf{C}^{-1}\mathbf{y} + \tilde{\Delta}^{-1}\mathbf{A}\hat{\bar{\beta}}_{GLS}\right), \end{aligned} \quad (46)$$

where

$$\tilde{\Delta} = \mathbf{I}_N \otimes \Delta.$$

When $E(\mathbf{u}_i \mathbf{u}_j') = \sigma_i^2 \mathbf{I}_T$ if $i = j$, and $\mathbf{0}$ otherwise, as assumed by Swamy (1971), we have

$$\widehat{\boldsymbol{\beta}}_i^* = \left[\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}_i}^{-1} + \boldsymbol{\Delta}^{-1} \right]^{-1} \left[\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}_i}^{-1} \widehat{\boldsymbol{\beta}}_i + \boldsymbol{\Delta}^{-1} \widehat{\boldsymbol{\beta}}_{GLS} \right], \quad i = 1, 2, \dots, N \quad (47)$$

where

$$\widehat{\boldsymbol{\beta}}_i = (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{y}_i, \quad \text{and} \quad \boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}_i} = V(\widehat{\boldsymbol{\beta}}_i) = \sigma_i^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1}. \quad (48)$$

The Bayes estimator (47), is identical to the Lindley and Smith (1972) estimator for a linear hierarchical model. This is to be expected since the Swamy type assumptions and the Lindley-Smith linear hierarchical model are formally equivalent.

The above estimator can also be written as

$$\widehat{\boldsymbol{\beta}}_i^* = \mathbf{H}_i \widehat{\boldsymbol{\beta}}_i + (\mathbf{I}_K - \mathbf{H}_i) \widehat{\boldsymbol{\beta}}_{GLS},$$

where

$$\mathbf{H}_i = \left[\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}_i}^{-1} + \boldsymbol{\Delta}^{-1} \right]^{-1} \boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}_i}^{-1} = \boldsymbol{\Delta} \left(\boldsymbol{\Delta} + \boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}_i} \right)^{-1}.$$

which shows that $\widehat{\boldsymbol{\beta}}_i^*$ is a weighted average of the OLS estimator, $\widehat{\boldsymbol{\beta}}_i$, and the Swamy estimator of $\bar{\boldsymbol{\beta}}$. Also, \mathbf{R}_i defined by (28) can be written as

$$\mathbf{R}_i = \left(\sum_{j=1}^N \mathbf{H}_j \right)^{-1} \mathbf{H}_i,$$

and hence

$$N^{-1} \sum_{i=1}^N \widehat{\boldsymbol{\beta}}_i^* = \sum_{i=1}^N \mathbf{R}_i \widehat{\boldsymbol{\beta}}_i = \widehat{\boldsymbol{\beta}}_{GLS},$$

namely the simple mean of the Bayes estimators (which could be viewed as the Bayes Mean Group estimator) is equal to the Swamy estimator of $\bar{\boldsymbol{\beta}}$.

Remark 1 *It is useful to put the random coefficients model in a Bayesian framework because many of the estimators based on the sampling approach can also be derived from the Bayes approach. For instance, as one can see from theorem 1(b) conditional on $\boldsymbol{\Delta}$ and \mathbf{C} , the Bayes estimator of $\bar{\boldsymbol{\gamma}}$ for the model (14) is identical to the GLS estimator of $\bar{\boldsymbol{\gamma}}$ (20). Furthermore, a Bayesian framework makes it clear the role of prior knowledge or ignorance about the parameter $\boldsymbol{\theta} = (\bar{\boldsymbol{\gamma}}, \boldsymbol{\alpha})$*

given \mathbf{y} . In a Bayesian approach, a different line is taken. The parameters $\boldsymbol{\theta}$ are treated as random variables and all probability statements are conditional. Ignorance about $\boldsymbol{\theta}$ would necessitate a specification of a diffuse prior to $\boldsymbol{\theta}$, which is typically specified as

$$p(\boldsymbol{\theta}) \propto \text{constant}.$$

On the other hand, information about $\boldsymbol{\theta}$ would necessitate a specification of an informative prior. The Swamy type random coefficients formulation of β_i having mean $\bar{\beta}$ and covariance Δ is equivalent to specifying an informative prior for the parameters β_i .

Remark 2 Typically, we use the expected value of an i.i.d. random variable as a predictor of the random variable. In panel data, we have two dimensions, a cross-sectional dimension and a time series dimension. Even though α_i is assumed independently distributed across i , once a particular α_i is drawn, it stays constant over time for the i th cross-sectional unit. Therefore, it makes sense to predict α_i . The Bayes predictor of α_i is different from the classical sampling approach predictor. For instance, for the Swamy type model the sampling approach predictor of $\beta_i = \bar{\beta} + \alpha_i$ defined by (23) is the least squares estimator (48). The Bayes predictor of β_i , given by (46) or (47), is a weighted average between the least squares estimator of β_i and the overall mean $\bar{\beta}$. In other words, the Bayes estimator of the individual coefficients β_i “shrinks” the estimate of β_i based on the information of the i th individual (48) towards the grand mean $\bar{\beta}$. An intuitive reason for doing so is because if the actual differences in β_i can be attributable to the work of chance mechanisms as postulated by de Finetti’s (1964) exchangeability assumption, information about β_i can be obtained by examining the behaviour of others in addition to those of the i th cross-sectional unit because the expected value of β_i is the same as β_j . When there are not many observations (i.e. T is small) with regard to the i th individual, information about β_i can be expanded by considering the responses of others. When T becomes large, more information about β_i becomes available and the weight gradually shifts towards the estimate based on the i th unit. As $T \rightarrow \infty$, the Bayes estimator approaches the least squares estimator $\hat{\beta}_i$.

Remark 3 The derivation of the posterior distribution and the Bayes estimators

$\bar{\gamma}$ and α of model (14) is based on known \mathbf{C} and Δ . When \mathbf{C} and Δ are unknown, in principle, we can first assign a joint prior of $\beta, \bar{\beta}, \mathbf{C}$ and Δ , and combine this with the likelihood function to obtain the joint posterior distribution. This distribution then has to be integrated with respect of \mathbf{C} and Δ . In practice, this is most complex to execute. Lindley and Smith (1972), therefore, suggest to approximate the posterior distribution of $\bar{\gamma}$ and α conditional on the modal value of Δ and \mathbf{C} . The modal estimates of Δ and \mathbf{C} may be found by supposing $\bar{\gamma}$ and α known, and then replacing $\bar{\gamma}$ and α in the results by their modes. The sequence of iterations typically starts with assumed values of Δ and \mathbf{C} to calculate the mode of $\bar{\gamma}$ and α , say $\hat{\bar{\gamma}}^{(1)}$ and $\hat{\alpha}^{(1)}$. Treating $\hat{\bar{\gamma}}^{(1)}$ and $\hat{\alpha}^{(1)}$ as known, we can find the mode for Δ and \mathbf{C} , say $\hat{\Delta}^{(1)}$ and $\mathbf{C}^{(1)}$. The $\hat{\Delta}^{(1)}$ and $\mathbf{C}^{(1)}$ are then used to find $\bar{\gamma}^{(2)}$ and $\hat{\alpha}^{(2)}$, and so on.

For the Swamy type model (6) and (7) under the assumption that Δ^{-1} has a Wishart distribution with p degrees of freedom and matrix \mathbf{R}^* , it is shown by Lindley and Smith (1972) that the mode estimator of Δ is

$$\hat{\Delta} = \left\{ \mathbf{R}^* + \sum_{i=1}^N (\hat{\beta}_i^* - \hat{\beta})(\hat{\beta}_i^* - \hat{\beta})' \right\} / (N + p - K - 2). \quad (49)$$

6 Dynamic Random Coefficients Models

Because of the inertia in human behaviour or institutional or technological rigidity, often a behavioural equation is specified with lagged dependent variable(s) appearing as regressor(s). We will consider a dynamic model of the form

$$y_{it} = \rho_i y_{i,t-1} + \mathbf{x}'_{it} \beta_i + u_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T, \quad (50)$$

where \mathbf{x}_{it} is a $K \times 1$ vector of exogenous variables, and the error term u_{it} is assumed to be independently, identically distributed over t with mean zero and variance σ_i^2 , and is independent across i . Let $\theta_i = (\rho_i, \beta_i)'$. We assume that θ_i is independently distributed across i with

$$E(\theta_i) = \bar{\theta} = (\bar{\rho}, \bar{\beta}')', \quad (51)$$

$$E[(\theta_i - \bar{\theta})(\theta_i - \bar{\theta})'] = \Delta. \quad (52)$$

Rewrite $\boldsymbol{\theta}_i = \bar{\boldsymbol{\theta}} + \boldsymbol{\alpha}_i$, (51) and (52) are equivalent to

$$E(\boldsymbol{\alpha}_i) = \mathbf{0}, \quad E(\boldsymbol{\alpha}_i \boldsymbol{\alpha}_j') = \begin{cases} \boldsymbol{\Delta} & \text{if } i = j, \\ \mathbf{0} & \text{if } i \neq j. \end{cases} \quad (53)$$

Although we may maintain the assumption (9) that $E(\boldsymbol{\alpha}_i \mathbf{x}_{it}') = \mathbf{0}$, we can no longer assume that $E(\boldsymbol{\alpha}_i y_{i,t-1}) = \mathbf{0}$. Through continuous substitutions, we have

$$\begin{aligned} y_{i,t-1} &= \sum_{j=0}^{\infty} (\bar{\rho} + \alpha_{i1})^j \mathbf{x}'_{i,t-j-1} (\bar{\boldsymbol{\beta}} + \boldsymbol{\alpha}_{i2}) \\ &\quad + \sum_{j=0}^{\infty} (\bar{\rho} + \alpha_{i1})^j u_{i,t-j-1}. \end{aligned} \quad (54)$$

It follows that $E(\boldsymbol{\alpha}_i y_{i,t-1}) \neq \mathbf{0}$.

The violation of the independence between the regressors and the individual effects $\boldsymbol{\alpha}_i$ implies that the pooled least squares regression of y_{it} on $y_{i,t-1}$, and \mathbf{x}_{it} will yield inconsistent estimates of $\bar{\boldsymbol{\theta}}$, even for T and N sufficiently large. Pesaran and Smith (1995) have noted that as $T \rightarrow \infty$, the least squares regression of y_{it} on $y_{i,t-1}$ and \mathbf{x}_{it} yields a consistent estimator of $\boldsymbol{\theta}_i$, $\hat{\boldsymbol{\theta}}_i$. They suggest a mean group estimator of $\bar{\boldsymbol{\theta}}$ by taking the average of $\hat{\boldsymbol{\theta}}_i$ across i ,

$$\hat{\boldsymbol{\theta}}_{MG} = \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\theta}}_i. \quad (55)$$

The mean group estimator is consistent when both N and $T \rightarrow \infty$. In finite T , $\hat{\boldsymbol{\theta}}_i$ for $\boldsymbol{\theta}_i$ is biased to the order of $1/T$. (Hurwicz (1950), Kiviet and Phillips (1993)) and the limited Monte Carlo appears to show that the mean group estimator can be severely biased when T is very small (Hsiao, Pesaran and Tahmiscioglu (1999)). However, under the assumption that y_{i0} are fixed and known and $\boldsymbol{\alpha}_i$ and u_{it} are independently normally distributed, as discussed in Section 5 we can implement the Bayes estimator of $\hat{\boldsymbol{\theta}}_i$ conditional on σ_i^2 and $\boldsymbol{\Delta}$,

$$\hat{\boldsymbol{\theta}}_B = \left\{ \sum_{i=1}^N [\sigma_i^2 (\mathbf{W}_i' \mathbf{W}_i)^{-1} + \boldsymbol{\Delta}]^{-1} \right\}^{-1} \sum_{i=1}^N [\sigma_i^2 (\mathbf{W}_i' \mathbf{W}_i)^{-1} + \boldsymbol{\Delta}] \hat{\boldsymbol{\theta}}_i, \quad (56)$$

where $\mathbf{W}_i = (\mathbf{y}_{i,-1}, \mathbf{X}_i)$ with $\mathbf{y}_{i,-1} = (y_{i0}, y_{i1}, \dots, y_{iT-1})'$. This Bayes estimator is a weighted average of the least squares estimator of individual units with the

weights being inversely proportional to individual variances. When $T \rightarrow \infty$, $N \rightarrow \infty$, and $N/T^{3/2} \rightarrow 0$, the Bayes estimator is asymptotically equivalent to the mean group estimator (55). (Hsiao, Pesaran and Tahmiscioglu (1999)).

In practice, the variance components, σ_i^2 and Δ are rarely known. The Monte Carlo studies conducted by Hsiao, Pesaran and Tahmiscioglu (1999) show that by following the approach of Lindley and Smith (1972) in assuming that the prior-distributions of σ_i^2 and Δ are independent and are distributed as

$$P(\Delta^{-1}, \sigma_1^2, \dots, \sigma_n^2) = W(\Delta^{-1} | (rR)^{-1}, r) \prod_{i=1}^N \sigma_i^{-2}, \quad (57)$$

yields a Bayes estimator almost as good as the Bayes estimator with known Δ and σ_i^2 , where $W(\cdot)$ represents the Wishart distribution with scale matrix, rR , and degrees of freedom r (e.g. Anderson (1985)).

The Hsiao, Pesaran and Tahmiscioglu (1999) Bayes estimator is derived under the assumption that the initial observation y_{i0} are fixed constants. As discussed in Anderson and Hsiao (1981, 1982), this assumption is clearly unjustifiable for a panel with finite T . However, contrary to the sampling approach where the correct modelling of initial observations is quite important, the Hsiao, Pesaran and Tahmiscioglu (1999) Bayesian approach appears to perform fairly well in the estimation of the mean coefficients for dynamic random coefficient models as demonstrated in their Monte Carlo studies.

Remark 4 *Model (50) has not imposed any constraint on the coefficient of the lag dependent variable, ρ_i . Often an investigator would like to impose the stability condition $|\rho_i| < 1$. One way to impose the stability condition on individual units would be to assume that ρ_i follows a Beta distribution on $(0,1)$. For a Bayes estimator under this assumption see Liu and Tiao (1980).*

7 Testing for Heterogeneity Under Weak Exogeneity

Given the importance of heterogeneity, it is very important to test for it. There are at least two different categories of tests available: direct tests of parameter

equality of the type used by Zellner (1962) in a SURE framework, or Hausman (1978) type tests of the difference between two estimators of $\boldsymbol{\theta}$ (or its subset). The former is generally applicable when N is relatively small and T sufficiently large. Here we shall examine the latter and assume that N and T are sufficiently large. The Hausman method can be used in cases where the two estimators are consistent under the null of homogeneity, whilst only one of them is efficient. Also, under the alternative hypothesis the two estimators converge to different values.

Denote the efficient estimator by subscript “e” and the inefficient but consistent estimator (under the alternative hypothesis) by the subscript “c”. Then we have

$$V(\widehat{\boldsymbol{\theta}}_c - \widehat{\boldsymbol{\theta}}_e) = V(\widehat{\boldsymbol{\theta}}_c) - V(\widehat{\boldsymbol{\theta}}_e). \quad (58)$$

This is the result used by Hausman (1978) where it is assumed that $\widehat{\boldsymbol{\theta}}_e$ is asymptotically the *most* efficient estimator. However, it is easily shown that (58) hold under a weaker requirement, namely when the (asymptotic) efficiency of $\widehat{\boldsymbol{\theta}}_e$ cannot be enhanced by the information contained in $\widehat{\boldsymbol{\theta}}_c$. Consider a third estimator $\widehat{\boldsymbol{\theta}}_*$, defined as a convex combination of $\widehat{\boldsymbol{\theta}}_c$ and $\widehat{\boldsymbol{\theta}}_e$

$$\mathbf{q}'\widehat{\boldsymbol{\theta}}_* = (1 - \delta)\mathbf{q}'\widehat{\boldsymbol{\theta}}_e + \delta\mathbf{q}'\widehat{\boldsymbol{\theta}}_c, \quad (59)$$

where \mathbf{q} is a vector of constants, and δ is a scalar in the range $0 \leq \delta \leq 1$. Since, by assumption, the asymptotic efficiency of $\widehat{\boldsymbol{\theta}}_e$ cannot be enhanced by the knowledge of $\widehat{\boldsymbol{\theta}}_c$, then it must be that $V(\mathbf{q}'\widehat{\boldsymbol{\theta}}_*) \geq V(\mathbf{q}'\widehat{\boldsymbol{\theta}}_e)$, and hence the value of δ that minimises $V(\mathbf{q}'\widehat{\boldsymbol{\theta}}_*)$, say δ^* , should be zero. However, using (59) directly, we have

$$\delta^* = \frac{\mathbf{q}'[V(\widehat{\boldsymbol{\theta}}_e) - Cov(\widehat{\boldsymbol{\theta}}_e, \widehat{\boldsymbol{\theta}}_c)]\mathbf{q}}{\mathbf{q}'V(\widehat{\boldsymbol{\theta}}_c - \widehat{\boldsymbol{\theta}}_e)\mathbf{q}} = 0, \quad (60)$$

and hence $\mathbf{q}'[V(\widehat{\boldsymbol{\theta}}_e) - Cov(\widehat{\boldsymbol{\theta}}_e, \widehat{\boldsymbol{\theta}}_c)]\mathbf{q} = 0$. But, if this result is to hold for an arbitrary vector \mathbf{q} , we must have

$$V(\widehat{\boldsymbol{\theta}}_e) = Cov(\widehat{\boldsymbol{\theta}}_e, \widehat{\boldsymbol{\theta}}_c). \quad (61)$$

Using this in

$$V(\widehat{\boldsymbol{\theta}}_c - \widehat{\boldsymbol{\theta}}_e) = V(\widehat{\boldsymbol{\theta}}_c) + V(\widehat{\boldsymbol{\theta}}_e) - 2 Cov(\widehat{\boldsymbol{\theta}}_e, \widehat{\boldsymbol{\theta}}_c),$$

yields (58) as desired.

In the context of testing for slope heterogeneity a number of different Hausman type tests can be used.

One possibility would be to compare the pooled estimator of $\bar{\theta}$, defined by²

$$\widehat{\theta}_{OLS} = \left(\sum_{i=1}^N \mathbf{w}_i' \mathbf{w}_i \right)^{-1} \sum_{i=1}^N \mathbf{w}_i' \mathbf{y}_i$$

with the mean group estimator $\widehat{\theta}_{MG}$, defined by (55). When the focus of attention is on the mean long run coefficients $\delta^* = E(\beta_i/(1 - \rho_i))$, as in Pesaran, Shin and Smith (1999) the heterogeneity test could be based directly on $\widehat{\delta}_{OLS}^* = \widehat{\beta}_{OLS}/(1 - \widehat{\rho}_{OLS})$ and $\widehat{\delta}_{MG}^* = N^{-1} \sum_{i=1}^N \widehat{\delta}_i^*$, where $\widehat{\delta}_i^* = \widehat{\beta}_i/(1 - \widehat{\rho}_i)$. Under the null of homogeneity the pooled and the mean group estimators are both consistent, although only the mean group estimator is consistent under the alternative hypothesis when lagged values of the dependent variables are included in the model.³

Under the full homogeneity assumption ($\theta_i = \theta$, $\sigma_i^2 = \sigma^2$), the asymptotic variance matrices of the pooled and the mean group estimators (for a fixed N and a large T) are given by

$$Cov(\sqrt{T}\widehat{\theta}_{OLS}) = \frac{\sigma^2}{N} \left(N^{-1} \sum_{i=1}^N \Psi_i \right)^{-1}, \quad (62)$$

and

$$Cov(\sqrt{T}\widehat{\theta}_{MG}) = \frac{\sigma^2}{N} \left(N^{-1} \sum_{i=1}^N \Psi_i^{-1} \right), \quad (63)$$

where $\Psi_i = p \lim_{T \rightarrow \infty} (\mathbf{W}_i' \mathbf{W}_i / T)$. Also we have

$$Cov(\sqrt{T}\widehat{\theta}_{OLS}, \sqrt{T}\widehat{\theta}_{MG}) = Cov(\sqrt{T}\widehat{\theta}_{OLS})$$

²Similar exercises can also be carried out using fixed or random effects estimators. But to keep the exposition simple here we focus on pooled estimators.

³Note that in the case of standard random coefficient models where all the regressors are strictly exogenous and the coefficients are distributed independently of the errors and the regressors, the pooled (OLS) estimators are consistent both under homogeneous and heterogeneous coefficients. As a result the Hausman test will not be applicable, in the sense that its application to the difference between pooled and non-pooled estimators will not result in a consistent test.

thus directly establishing that

$$Cov \left[\sqrt{T} \left(\widehat{\boldsymbol{\theta}}_{MG} - \widehat{\boldsymbol{\theta}}_{OLS} \right) \right] = \frac{\sigma^2}{N} \left\{ \left(N^{-1} \sum_{i=1}^N \boldsymbol{\Psi}_i^{-1} \right) - \left(N^{-1} \sum_{i=1}^N \boldsymbol{\Psi}_i \right)^{-1} \right\},$$

which is a positive definite matrix, assuming that $\boldsymbol{\Psi}_i \neq \boldsymbol{\Psi}_j$, for some i and j .⁴ This condition is generally satisfied when the model contains regressors with heterogeneous variances. The above results suggest the following statistic for testing the homogeneity hypothesis:

$$h = N T \left(\widehat{\boldsymbol{\theta}}_{MG} - \widehat{\boldsymbol{\theta}}_{OLS} \right)' \widehat{\mathbf{V}}^{-1} \left(\widehat{\boldsymbol{\theta}}_{MG} - \widehat{\boldsymbol{\theta}}_{OLS} \right),$$

where

$$\widehat{\mathbf{V}} = \widehat{\sigma}_{MG}^2 \left\{ N^{-1} \sum_{i=1}^N \left(\frac{\mathbf{W}'_i \mathbf{W}_i}{T} \right)^{-1} - \left(N^{-1} \sum_{i=1}^N \frac{\mathbf{W}'_i \mathbf{W}_i}{T} \right)^{-1} \right\}. \quad (64)$$

and $\widehat{\sigma}_{MG}^2 = \frac{1}{N} \sum_{i=1}^N \widehat{\sigma}_i^2$. In computing h , one could also equally use $\widehat{\sigma}_{OLS}^2$ instead of $\widehat{\sigma}_{MG}^2$. Under the null hypothesis

$$H_0 : \rho_i = \rho, \boldsymbol{\beta}_i = \boldsymbol{\beta}, \text{ and } \sigma_i^2 = \sigma^2, \text{ for all } i,$$

and for N and T sufficiently large we have

$$h \stackrel{a}{\sim} \chi_{K+1}^2.$$

When the focus of the analysis is on the long run coefficients we first note that⁵

$$\widehat{\boldsymbol{\delta}}_{OLS}^* - \boldsymbol{\delta}^* = \frac{(\widehat{\rho}_{OLS} - \rho) \boldsymbol{\delta}^* + (\widehat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta})}{(1 - \widehat{\rho}_{OLS})}.$$

Therefore, under the homogeneity hypothesis, we have, for a large T

$$Cov \left(\sqrt{T} \widehat{\boldsymbol{\delta}}_{OLS}^* \right) = \frac{\sigma^2}{N(1 - \rho)^2} \mathbf{D} \left(N^{-1} \sum_{i=1}^N \boldsymbol{\Psi}_i \right)^{-1} \mathbf{D}'. \quad (65)$$

Similarly,

$$Cov \left(\sqrt{T} \widehat{\boldsymbol{\theta}}_{MG}^* \right) = \frac{\sigma^2}{N(1 - \rho)^2} \mathbf{D} \left(N^{-1} \sum_{i=1}^N \boldsymbol{\Psi}_i^{-1} \right) \mathbf{D}'. \quad (66)$$

⁴For a proof see the Appendix in Pesaran, Smith and Im (1996).

⁵Recall that under homogeneity hypothesis $\boldsymbol{\delta}^* = \boldsymbol{\beta}/(1 - \rho)$ and $\widehat{\boldsymbol{\delta}}_{OLS}^* = \widehat{\boldsymbol{\beta}}_{OLS}/(1 - \widehat{\rho}_{OLS})$.

where $\mathbf{D} = (\boldsymbol{\delta}^*, \mathbf{I}_K)$ is a $K \times (K + 1)$ matrix. To estimate (65), and (66), the unknown parameters σ^2, ρ , and $\boldsymbol{\theta}$ could be estimated either from pooled (fixed effects) or mean group estimators. Using the mean group estimators, the test of the homogeneity or the long run coefficients can then be based on the following Hausman-type statistic:

$$h_{\delta^*} = NT(1 - \hat{\rho}_{MG})^2 \left(\hat{\boldsymbol{\delta}}_{MG}^* - \hat{\boldsymbol{\delta}}_{OLS}^* \right)' \left(\hat{\mathbf{D}}_{MG} \hat{\mathbf{V}} \hat{\mathbf{D}}_{MG}' \right)^{-1} \left(\hat{\boldsymbol{\delta}}_{MG}^* - \hat{\boldsymbol{\delta}}_{OLS}^* \right),$$

where $\hat{\mathbf{D}}_{MG} = (\hat{\boldsymbol{\delta}}_{MG}^*, \mathbf{I}_K)$, and $\hat{\mathbf{V}}$ is given by (64). In general $\hat{\mathbf{D}}_{MG} \hat{\mathbf{V}} \hat{\mathbf{D}}_{MG}'$ is of full rank. Under the null hypothesis, for large N and T , $h_{\delta^*} \sim \chi_K^2$.

8 A Random Coefficient Simultaneous Equation System

The generalisation of a single equation random coefficients model to a simultaneous equation system raises complicated issues of identification and estimation. To show this let us consider a system of G equations

$$\mathbf{Y}_i \mathbf{B}_i + \mathbf{X}_i \boldsymbol{\Gamma}_i = \mathbf{U}_i, \quad i = 1, \dots, N, \quad (67)$$

where \mathbf{Y}_i and \mathbf{X}_i are the $T \times G$ and $T \times K$ matrices of endogenous and exogenous variables, respectively, \mathbf{U}_i is the $T \times G$ matrices of errors, \mathbf{B}_i and $\boldsymbol{\Gamma}_i$ are the $G \times G$ and $K \times G$ matrix of the coefficients of the endogenous variables and exogenous variables, respectively. The reduced form, then, is of the form

$$\begin{aligned} \mathbf{Y}_i &= -\mathbf{X}_i \boldsymbol{\Gamma}_i \mathbf{B}_i^{-1} + \mathbf{U}_i \mathbf{B}_i^{-1}, \\ &= \mathbf{X}_i \boldsymbol{\Pi}_i + \mathbf{V}_i, \end{aligned} \quad (68)$$

where

$$\boldsymbol{\Pi}_i = -\boldsymbol{\Gamma}_i \mathbf{B}_i^{-1}, \quad (69)$$

$$\mathbf{V}_i = \mathbf{U}_i \mathbf{B}_i^{-1}. \quad (70)$$

Suppose that

$$\mathbf{B}_i = \bar{\mathbf{B}} + \boldsymbol{\xi}_i, \quad (71)$$

$$\boldsymbol{\Gamma}_i = \bar{\boldsymbol{\Gamma}} + \boldsymbol{\alpha}_i, \quad (72)$$

where $\boldsymbol{\xi}_i$ and $\boldsymbol{\alpha}_i$ are $G \times G$ and $G \times K$ matrices of random variables independently distributed over i with means $\mathbf{0}$ and covariances $\boldsymbol{\Phi}$ and $\boldsymbol{\Delta}$, defined by $\boldsymbol{\Phi} = E[(\text{vec } \boldsymbol{\xi}_i)(\text{vec } \boldsymbol{\xi}_i)']$ and $\boldsymbol{\Delta} = E[(\text{vec } \boldsymbol{\alpha}_i)(\text{vec } \boldsymbol{\alpha}_i)']$. Then

$$\begin{aligned} E(\boldsymbol{\Pi}_i) &= -E[(\bar{\boldsymbol{\Gamma}} + \boldsymbol{\alpha}_i)(\bar{\mathbf{B}} + \boldsymbol{\xi}_i)^{-1}], \\ &\neq \bar{\boldsymbol{\Gamma}} \bar{\mathbf{B}}^{-1}. \end{aligned} \quad (73)$$

In other words, identification conditions of structural parameters cannot be derived by assuming that when sample size approaches infinity, $\hat{\boldsymbol{\Pi}}$ will converge to $\bar{\boldsymbol{\Gamma}} \bar{\mathbf{B}}^{-1}$. In fact the assumption of (71) raises intractable difficulties at the levels of identification and estimation.

Kelejian (1974) has studied the problem of identification under (71) and (72). His results imply that any feedback between the endogenous variables must be avoided and that identifiability and interdependence exclude each other (also see Raj and Ullah (1981)). In other words, for any one equation we may treat all the other variables as predetermined. Therefore, for ease of analysis, instead of assuming (71), we shall assume that

$$\mathbf{B}_i = \bar{\mathbf{B}}, \quad \forall i, \quad (74)$$

where $\bar{\mathbf{B}}$ is a non-singular matrix with fixed elements.

The combination of (72) and (74) amounts to assuming a random coefficients reduced form of (51), where $\boldsymbol{\Pi}_i = -\boldsymbol{\Gamma}_i \bar{\mathbf{B}}^{-1} = -(\bar{\boldsymbol{\Gamma}} + \boldsymbol{\alpha}_i) \bar{\mathbf{B}}^{-1}$, and

$$E(\boldsymbol{\Pi}_i) = -\bar{\boldsymbol{\Gamma}} \bar{\mathbf{B}}^{-1}, \quad (75)$$

$$\text{Cov}(\boldsymbol{\Pi}_i) = [\bar{\mathbf{B}}^{-1'} \otimes \mathbf{I}_k] \boldsymbol{\Delta} [\bar{\mathbf{B}}^{-1} \otimes \mathbf{I}_k] = \boldsymbol{\Delta}^*. \quad (76)$$

Assume that \mathbf{U}_i are independently distributed over time but are contemporaneously correlated, then

$$\text{Cov}(\mathbf{U}_i) = E[\text{vec}(\mathbf{U}_i) \text{vec}(\mathbf{U}_i)'] = \mathbf{C}_i \otimes \mathbf{I}_T. \quad (77)$$

Furthermore, we assume that \mathbf{U}_i and $\boldsymbol{\alpha}_i$ are mutually independent and are independent of \mathbf{X}_i . Then the reduced form (68) can be written as

$$\mathbf{Y}_i = \mathbf{X}_i \bar{\boldsymbol{\Pi}} + \mathbf{V}_i^*, \quad (78)$$

where $\mathbf{V}_i^* = -\mathbf{X}_i\boldsymbol{\alpha}_i\bar{\mathbf{B}}^{-1} + \mathbf{U}_i\bar{\mathbf{B}}^{-1}$ and $E(\mathbf{V}_i^*) = \mathbf{0}$,

$$\begin{aligned} Cov(\mathbf{V}_i^*) &= \bar{\mathbf{B}}^{-1'}\mathbf{C}_i\bar{\mathbf{B}}^{-1} \otimes \mathbf{I}_T + (\mathbf{I}_G \otimes \mathbf{X}_i)\boldsymbol{\Delta}^*(\mathbf{I}_G \otimes \mathbf{X}_i') \\ &= \mathbf{Q}_i, \quad i = 1, \dots, N. \end{aligned} \quad (79)$$

The GLS estimator of $\bar{\boldsymbol{\Pi}}$ is then equal to (Balestra and Negassi (1990))

$$vec(\widehat{\bar{\boldsymbol{\Pi}}}_{GLS}) = \left[\sum_{i=1}^N \tilde{\mathbf{R}}_i^{-1} \right]^{-1} \left[\sum_{i=1}^N \tilde{\mathbf{R}}_i^{-1} vec \hat{\boldsymbol{\Pi}}_i \right], \quad (80)$$

where

$$\tilde{\mathbf{R}}_i = \mathbf{Q}_i \otimes (\mathbf{X}_i'\mathbf{X}_i)^{-1} + \boldsymbol{\Delta}^*, \quad (81)$$

$$\hat{\boldsymbol{\Pi}}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{Y}_i. \quad (82)$$

If $\bar{\mathbf{B}}^{-1'}\mathbf{C}_i\bar{\mathbf{B}}^{-1}$ and $\boldsymbol{\Delta}^*$ are unknown, a two-step GLS procedure can be applied. In the first step, we estimate $\bar{\mathbf{B}}^{-1'}\mathbf{C}_i\bar{\mathbf{B}}^{-1}$ and $\boldsymbol{\Delta}^*$ by

$$\begin{aligned} \bar{\mathbf{B}}^{-1'}\widehat{\mathbf{C}}_i\bar{\mathbf{B}}^{-1} &= \frac{1}{T-K}\tilde{\mathbf{V}}_i^{*'}\tilde{\mathbf{V}}_i^*, \quad \tilde{\mathbf{V}}_i^* = \mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\Pi}}_i, \\ vec(\boldsymbol{\Delta}^*) &= \frac{1}{N-1} \sum_{i=1}^N [vec(\hat{\boldsymbol{\Pi}}_i - \bar{\boldsymbol{\Pi}})][vec(\hat{\boldsymbol{\Pi}}_i - \bar{\boldsymbol{\Pi}})]', \\ \bar{\boldsymbol{\Pi}} &= \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\Pi}}_i. \end{aligned} \quad (83)$$

In the second step, we estimate $\bar{\boldsymbol{\Pi}}$ using (80) by substituting $\widehat{\tilde{\mathbf{R}}}_i$ for $\tilde{\mathbf{R}}_i$.

If our interest is in the structural form parameters $\bar{\mathbf{B}}$ and $\bar{\boldsymbol{\Gamma}}$, we can either solve for $\bar{\mathbf{B}}$ and $\bar{\boldsymbol{\Gamma}}$ from the reduced form estimate $\widehat{\bar{\boldsymbol{\Pi}}}$, or we can estimate them directly using instrumental variables method. Rewrite the first equation of the structural form in the following way,

$$\begin{aligned} \mathbf{y}_{i1} &= \mathbf{Y}_{i1}\bar{\boldsymbol{\beta}}_1^* + \mathbf{X}_{i1}\boldsymbol{\gamma}_{i1}^* + \mathbf{u}_{i1}, \\ &= \mathbf{Z}_{i1}\boldsymbol{\delta}_1 + \mathbf{v}_{i1}^*, \quad i = 1, \dots, N, \end{aligned} \quad (84)$$

where \mathbf{y}_{i1} is the $T \times 1$ vector of the first endogenous variables and \mathbf{Y}_{i1} is the $T \times g$ matrix of the other endogenous variables appearing in the first equation $g \leq G - 1$, \mathbf{X}_{i1} is the $T \times k^*$ matrix of included exogenous variables $k^* \leq K$,

and $\bar{\beta}_1^*$ and γ_{i1}^* are $g \times 1$ and $k^* \times 1$ vectors of coefficients, respectively with $\gamma_{i1}^* = [\bar{\gamma}_{i1}^* + \alpha_{i1}^*]$, and $\mathbf{Z}_{i1} = [\mathbf{Y}_{i1}, \mathbf{X}_{i1}]$, $\delta_1' = [\bar{\beta}_1^{*'}, \bar{\gamma}_1^{*'}]$, $\mathbf{v}_{i1}^* = \mathbf{u}_{i1} + \mathbf{X}_{i1}\alpha_{i1}^*$. Balestra and Negassi (1990) suggest the following instrumental variables estimator

$$\hat{\delta}_1 = \left[\sum_{i=1}^N \mathbf{Z}'_{i1} \mathbf{F}_i \mathbf{Z}_{i1} \right]^{-1} \left[\sum_{i=1}^N \mathbf{Z}'_{i1} \mathbf{F}_i \hat{\delta}_{i1} \right], \quad (85)$$

where

$$\hat{\delta}_{i1} = [\mathbf{Z}'_{i1} \mathbf{X}_i (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{Z}_{i1}]^{-1} \mathbf{Z}'_{i1} \mathbf{X}_i (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{y}_{i1}, \quad (86)$$

and

$$\begin{aligned} \mathbf{F}_i &= \mathbf{X}_i (\mathbf{X}'_{i1} \Delta_1 \mathbf{X}'_{i1} + \sigma_1^2 \mathbf{I}_{k^*})^{-1} \mathbf{X}'_i, \\ \Delta_1 &= E(\alpha_{i1} \alpha'_{i1}). \end{aligned} \quad (87)$$

One can also derive the Bayes solutions for a simultaneous equations system of the form (67), (72) and (74) using a method analogous to that of section 4. Considering one equation of (68) at a time, the results of section 4 can be applied straightforwardly. The similar results for the system of (68) can also be derived if the prior restrictions on $\bar{\Pi}$ are ignored. Of course, restricted reduced form estimators can also be derived. The computation, though, can be laborious.

The results of section 4 can also be used to derive a Bayes estimator for the structural form (84) based on a limited information approach. Let

$$\mathbf{Y}_{i1} = \hat{\mathbf{Y}}_{i1} + \hat{\mathbf{V}}_{i1}, \quad (88)$$

where $\hat{\mathbf{Y}}_{i1} = \mathbf{X}_i \hat{\Pi}_{i1}$, and $\hat{\Pi}_{i1} = (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{Y}_{i1}$. Substituting $\hat{\mathbf{Y}}_{i1}$ for \mathbf{Y}_{i1} in (84), we have

$$\mathbf{y}_{i1} = \hat{\mathbf{Y}}_{i1} \bar{\beta}_1^* + \mathbf{X}_{i1} \gamma_{i1}^* + \boldsymbol{\eta}_{i1}, \quad (89)$$

where $\boldsymbol{\eta}_{i1} = \mathbf{u}_{i1} + \hat{\mathbf{V}}_{i1} \bar{\beta}_1^*$. Conditioning on $\hat{\Pi}_{i1}$, we can treat $\hat{\mathbf{Y}}_{i1}$ and \mathbf{X}_{i1} as the set of exogenous variables. Equation (89) is of the form of the mixed fixed and random coefficients model (14) and the Bayes estimators of $\bar{\beta}_1^*$, $\bar{\gamma}_1^*$ and γ_{i1}^* are given in section 4 (for detail see Hsiao, Appelbe and Dineen (1991)). Of course, one should keep in mind that now the Bayes estimator is the conditional posterior mean given the estimated Π_{i1} .

9 Random Coefficient Models with Cross-Section Dependence

In principle, the random coefficient model (14) can be easily adapted to allow for dependence across the error terms, u_{it} , $i = 1, 2, \dots, N$. But, without plausible restrictions on the error covariances the number of unknown parameters of the model increases at the rate of N^2 , which would be manageable only when N is relatively small (typically 10 or less). To deal with the problem of cross section dependence when N is large a number of different approaches have been advanced in the literature. In the case of spatial panels where a natural distance measure (or an immutable ordering of cross section units) is available the dependence is tested and modelled with “spatial lags”, using techniques familiar from the time series literature. Anselin (2001) provides a recent survey of the literature on spatial econometrics. A number of studies have also used measures such as trade or capital flows to capture economic distance, as in Lee and Pesaran (1993), Conley and Topa (2002) and Pesaran, Schuermann and Weiner (2004).

But, in the absence of suitable distance measures or natural orderings of the cross section units a number of investigators have attempted to model the cross section dependence using single or multiple factor residual models where u_{it} is specified in terms of a finite number of common factors.⁶ A convenient parameterization is given by

$$u_{it} = \frac{\sigma_i}{\sqrt{1 + \boldsymbol{\delta}'_i \boldsymbol{\delta}_i}} (\boldsymbol{\delta}'_i \mathbf{f}_t + \varepsilon_{it}), \quad (90)$$

where $\boldsymbol{\delta}_i$ is a $s \times 1$ vector of individual-specific factor loadings, \mathbf{f}_t is an $s \times 1$ vector of unobserved (latent) factors, and ε_{it} is an idiosyncratic error assumed to be distributed independently across i , the unobserved factors, \mathbf{f}_t , and the observed regressors, \mathbf{x}_{it} , with mean zero and a unit variance. Since the common factors are unobserved, without loss of generality we also assume that $\mathbf{f}_t \sim (\mathbf{0}, \mathbf{I}_s)$.

Under the above set up, and conditional on a *given* set of factor loadings, the cross-correlations of the errors are given by

$$\rho_{ij} = \rho_{ji} = \frac{\boldsymbol{\delta}'_i \boldsymbol{\delta}_j}{(1 + \boldsymbol{\delta}'_i \boldsymbol{\delta}_i)^{1/2} (1 + \boldsymbol{\delta}'_j \boldsymbol{\delta}_j)^{1/2}}. \quad (91)$$

⁶A general test of cross section dependence is also proposed by Pesaran (2004).

Complicated covariance structures can be accommodated by the residual factor formulation through differences across factor loadings and by using a sufficiently large number of factors. A random coefficient specification can also be assumed for the factor loadings:

$$\boldsymbol{\delta}_i = \bar{\boldsymbol{\delta}} + \boldsymbol{\zeta}_i, \quad (92)$$

where $\bar{\boldsymbol{\delta}}$ is a vector of fixed constants

$$\begin{aligned} E(\boldsymbol{\zeta}_i) &= \mathbf{0}, E(\boldsymbol{\zeta}_i \mathbf{f}_t') = \mathbf{0}, \\ E(\boldsymbol{\zeta}_i \mathbf{x}_{it}') &= \mathbf{0}, E(\boldsymbol{\zeta}_i \boldsymbol{\alpha}_i') = \mathbf{0}, \\ E(\boldsymbol{\zeta}_i \boldsymbol{\zeta}_i') &= \begin{cases} \boldsymbol{\Delta}_\zeta, & \text{if } i = j, \\ \mathbf{0}, & \text{if } i \neq j, \end{cases} \end{aligned} \quad (93)$$

and $\boldsymbol{\Delta}_\zeta$ is a non-negative definite matrix. The average degree of cross dependence, defined by $E(\rho_{ij})$ for $i \neq j$ is governed by $\bar{\boldsymbol{\delta}}$ and the distribution of $\boldsymbol{\zeta}_i$. The average cross section dependence will be zero if $\bar{\boldsymbol{\delta}} = \mathbf{0}$, and $\boldsymbol{\zeta}_i$ is symmetrically distributed. Typically one would expect $\bar{\boldsymbol{\delta}} \neq \mathbf{0}$.

Examples of studies that have used the residual factor structure to model cross section dependence include Holtz-Eakin, Newey, and Rosen (1988), Ahn, Lee and Schmidt (2001), Robertson and Symons (2000), Bai and Ng (2002), Coakley, Fuertes and Smith (2002), Pesaran (2002, 2003), Phillips and Sul (2002), Moon and Perron (2003), and Moon, Perron and Phillips (2003). The studies by Holtz-Eakin et al. and Ahn et al. focus on single factor residual models and allow for time-varying individual effects in the case of panels with homogeneous slopes where T is fixed and $N \rightarrow \infty$. Robertson and Symons consider a random coefficient multi-factor residual model where the factors are distributed independently of regressors, \mathbf{x}_{it} , and argue that the maximum likelihood method would still be applicable even when $N > T$. Phillips and Sul (2002) suggest using SURE-GLS techniques combined with median unbiased estimation in the case of first order autoregressive panels. Coakley, Fuertes and Smith propose a principal components approach which is shown by Pesaran (2002) to be consistent only when the factors and the included regressors are either asymptotically uncorrelated or are perfectly correlated. In the more general case Pesaran (2002) shows that consistent estimation of the random coefficient models with a residual factor structure can be achieved (under certain regularity conditions) by augmenting

the observed regressors with the cross section averages of the dependent variable and individual-specific regressors, namely

$$\bar{y}_t = \sum_{j=1}^N w_j y_{jt}, \text{ and } \bar{\mathbf{x}}_{it} = \sum_{j=1}^N w_j \mathbf{x}_{jt}, \quad (94)$$

for any set of weights such that

$$w_i = O\left(\frac{1}{N}\right), \quad \sum_{i=1}^N |w_i| < K < \infty.$$

An obvious example of such a weighting scheme is $w_i = 1/N$.

Bai and Ng (2002), Phillips and Sul (2002), Moon and Perron (2003), and Moon, Perron and Phillips (2003), and Pesaran (2003) use residual factor models to allow for cross section dependence in testing for unit roots in Panels.

10 Concluding Remarks

When the included conditional variables together with the conventional variable intercept or error components (e.g. Hsiao (2003, ch.3)) cannot completely captured systematic differences across cross-sectional units and/or over time, and the possibility of adding additional conditional variables is not an option, either due to data unavailability or the desire to keep the model simple, there is very little alternative but to allow the slope coefficients to vary across cross-section units or over time. If we treat all these coefficients as fixed and different, there is no particular reason to pool the data, except for some efficiency gain in a Zellner's (1962) seemingly unrelated regression framework. Random coefficients models appear to be an attractive middle ground between the implausible assumption of homogeneity across cross-sectional units or over time and the infeasibility of treating them all differently, in the sense of being draws from different probability distributions. Other intermediate formulations could also be considered. For example, as argued by Pesaran, Shin and Smith (1999), in the context of dynamic models it would be plausible to impose the homogeneity hypothesis on the long-run coefficients but let the short-run dynamics to vary freely across the cross-section units. In this Chapter various formulations are surveyed and their

implications discussed. Our review has been largely confined to linear panel data models with stationary regressors. The analysis of random coefficient models with unit roots and cointegration is reviewed in Breitung and Pesaran (2004, under preparation) in this volume. Parameter heterogeneity in non-linear panel data models poses fundamentally new problems and needs to be considered on a case-by-case basis.

Appendix A: Proof of Theorem 1

To prove part (a) of the theorem, we write (41) in the form of (19) and (17). Putting $\mathbf{u} \sim \mathbf{N}(\mathbf{0}, \mathbf{C})$ and $\boldsymbol{\alpha} \sim \mathbf{N}(\mathbf{0}, \mathbf{I}_N \otimes \boldsymbol{\Delta})$ together with (17), the result follows.

To prove (b), we use Bayes's theorem, that is

$$p(\bar{\boldsymbol{\gamma}}|\mathbf{y}) \propto p(\mathbf{y}|\bar{\boldsymbol{\gamma}})p(\bar{\boldsymbol{\gamma}}), \quad (95)$$

where $p(\mathbf{y}|\bar{\boldsymbol{\gamma}})$ follows from (42) and $p(\bar{\boldsymbol{\gamma}})$ is given by (39). The product on the right hand side of (95) is proportional to $\exp\{-\frac{1}{2}Q\}$, where Q is given by

$$\begin{aligned} Q &= (\mathbf{y} - \bar{\mathbf{Z}}\bar{\boldsymbol{\gamma}})'[\mathbf{C} + \mathbf{W}(\mathbf{I}_N \otimes \boldsymbol{\Delta})\mathbf{W}']^{-1}(\mathbf{y} - \bar{\mathbf{Z}}\bar{\boldsymbol{\gamma}}) \\ &= (\bar{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}})'\mathbf{D}^{-1}(\bar{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}}) + \mathbf{y}'\{\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\bar{\mathbf{Z}}[\bar{\mathbf{Z}}'\mathbf{D}\bar{\mathbf{Z}}]^{-1}\bar{\mathbf{Z}}'\boldsymbol{\Omega}^{-1}\}\mathbf{y}. \end{aligned} \quad (96)$$

The second term on the right hand side of (96) is a constant as far as the distribution of $\bar{\boldsymbol{\gamma}}$ is concerned, and the remainder of the expression demonstrates the truth of (b).

To prove (c), we use the relations

$$\begin{aligned} p(\boldsymbol{\alpha}|\mathbf{y}) &= \int p(\boldsymbol{\alpha}, \bar{\boldsymbol{\gamma}}|\mathbf{y})d\bar{\boldsymbol{\gamma}} \\ &= \int [p(\bar{\boldsymbol{\gamma}}|\mathbf{y}, \boldsymbol{\alpha})d\bar{\boldsymbol{\gamma}}]p(\boldsymbol{\alpha}|\mathbf{y}) \end{aligned} \quad (97)$$

and

$$\begin{aligned} p(\boldsymbol{\alpha}, \bar{\boldsymbol{\gamma}}|\mathbf{y}) &\propto p(\mathbf{y}|\boldsymbol{\alpha}, \bar{\boldsymbol{\gamma}})p(\boldsymbol{\alpha}, \bar{\boldsymbol{\gamma}}) \\ &= p(\mathbf{y}|\boldsymbol{\alpha}, \bar{\boldsymbol{\gamma}})p(\boldsymbol{\alpha}) \cdot p(\bar{\boldsymbol{\gamma}}). \end{aligned} \quad (98)$$

Under (38) - (40), the right hand side of (98) is proportional to $\exp\{-\frac{1}{2}Q^*\}$, where Q^* is given by

$$\begin{aligned} Q^* &= (\mathbf{y} - \bar{\mathbf{Z}}\bar{\boldsymbol{\gamma}} - \mathbf{W}\boldsymbol{\alpha})'\mathbf{C}^{-1}(\mathbf{y} - \bar{\mathbf{Z}}\bar{\boldsymbol{\gamma}} - \mathbf{W}\boldsymbol{\alpha}) + \boldsymbol{\alpha}'(\mathbf{I}_N \otimes \boldsymbol{\Delta}^{-1})\boldsymbol{\alpha} \\ &= \mathbf{y}'\mathbf{C}^{-1}\mathbf{y} + \bar{\boldsymbol{\gamma}}'\bar{\mathbf{Z}}'\mathbf{C}^{-1}\bar{\mathbf{Z}}\bar{\boldsymbol{\gamma}} + \boldsymbol{\alpha}'\mathbf{W}'\mathbf{C}^{-1}\mathbf{W}\boldsymbol{\alpha} \\ &\quad - 2\bar{\boldsymbol{\gamma}}'\bar{\mathbf{Z}}'\mathbf{C}^{-1}\mathbf{y} - 2\boldsymbol{\alpha}'\mathbf{W}'\mathbf{C}^{-1}\mathbf{y} + 2\bar{\boldsymbol{\gamma}}'\bar{\mathbf{Z}}'\mathbf{C}^{-1}\mathbf{W}\boldsymbol{\alpha} + \boldsymbol{\alpha}'(\mathbf{I}_N \otimes \boldsymbol{\Delta}^{-1})\boldsymbol{\alpha} \\ &= Q_1^* + Q_2^* + Q_3^*, \end{aligned} \quad (99)$$

with

$$Q_1^* = \{\bar{\gamma} - (\bar{\mathbf{Z}}'\mathbf{C}^{-1}\bar{\mathbf{Z}})^{-1}[\bar{\mathbf{Z}}'\mathbf{C}^{-1}(\mathbf{y} - \mathbf{W}\boldsymbol{\alpha})]\}'(\bar{\mathbf{Z}}'\mathbf{C}^{-1}\bar{\mathbf{Z}}) \cdot \{\bar{\gamma} - (\bar{\mathbf{Z}}'\mathbf{C}^{-1}\bar{\mathbf{Z}})^{-1}[\bar{\mathbf{Z}}'\mathbf{C}^{-1}(\mathbf{y} - \mathbf{W}\boldsymbol{\alpha})]\}, \quad (100)$$

$$Q_2^* = \{\boldsymbol{\alpha} - \tilde{\mathbf{D}}\mathbf{W}'[\mathbf{C}^{-1} - \mathbf{C}^{-1}\bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\mathbf{C}^{-1}\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{C}^{-1}]\mathbf{y}\}'\tilde{\mathbf{D}}^{-1} \cdot \{\boldsymbol{\alpha} - \tilde{\mathbf{D}}\mathbf{W}'[\mathbf{C}^{-1} - \mathbf{C}^{-1}\bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\mathbf{C}^{-1}\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{C}^{-1}]\mathbf{y}\} \quad (101)$$

and

$$Q_3^* = \mathbf{y}'\{\mathbf{C}^{-1} - \mathbf{C}^{-1}\bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\mathbf{C}^{-1}\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{C}^{-1} - [\mathbf{C}^{-1} - \mathbf{C}^{-1}\bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\mathbf{C}^{-1}\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{C}^{-1}] \cdot \mathbf{W}\tilde{\mathbf{D}}^{-1}\mathbf{W}'[\mathbf{C}^{-1} - \mathbf{C}^{-1}\bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\mathbf{C}^{-1}\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{C}^{-1}]\}\mathbf{y}. \quad (102)$$

As far as the distribution of $p(\boldsymbol{\alpha}, \bar{\gamma}|\mathbf{y})$ is concerned, Q_3^* is a constant. The conditional distribution of $\bar{\gamma}$ given \mathbf{y} and $\boldsymbol{\alpha}$ is proportional to $\exp\{-\frac{1}{2}Q_1^*\}$, which integrates to 1. Therefore, the marginal distribution of $\boldsymbol{\alpha}$ given \mathbf{y} is proportional to $\exp\{-\frac{1}{2}Q_2^*\}$, which demonstrates (c).

Substituting (23) - (26) into (42) we obtain the Bayes solutions for the Swamy type random coefficients model: (i) the distribution of $\bar{\boldsymbol{\beta}}$ given \mathbf{y} is $N(\hat{\bar{\boldsymbol{\beta}}}, \mathbf{D})$, and (ii) the distribution of $\boldsymbol{\alpha}$ given \mathbf{y} is normal with mean

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \{\mathbf{X}'[\mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{X}\mathbf{A}(\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}]\mathbf{X} + (\mathbf{I}_N \otimes \boldsymbol{\Delta}^{-1})\}^{-1} \\ &\quad \cdot \{\mathbf{X}'[\mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{X}\mathbf{A}(\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}]\mathbf{y}\} \\ &= \tilde{\mathbf{D}}\{\mathbf{X}'[\mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{X}\mathbf{A}(\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}]\mathbf{y}\}, \end{aligned} \quad (103)$$

and covariance

$$\tilde{\mathbf{D}} = \{\mathbf{X}'[\mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{X}\mathbf{A}(\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}]\mathbf{X} + (\mathbf{I}_N \otimes \boldsymbol{\Delta}^{-1})\}^{-1}. \quad (104)$$

Letting $\tilde{\boldsymbol{\Delta}} = \mathbf{I}_N \otimes \boldsymbol{\Delta}$ and repeatedly using the identity (30) we can write (104) in the form

$$\begin{aligned} \tilde{\mathbf{D}} &= [\mathbf{X}'\mathbf{C}^{-1}\mathbf{X} + \tilde{\boldsymbol{\Delta}}^{-1}]^{-1}\{\mathbf{I} - \mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\mathbf{A}[\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X} + \tilde{\boldsymbol{\Delta}}^{-1})^{-1}\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\mathbf{A} \\ &\quad - \mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\mathbf{A}]^{-1}\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}[\mathbf{X}'\mathbf{C}^{-1}\mathbf{X} + \tilde{\boldsymbol{\Delta}}^{-1}]^{-1}\} \\ &= [\mathbf{X}'\mathbf{C}^{-1}\mathbf{X} + \tilde{\boldsymbol{\Delta}}^{-1}]^{-1}\{\mathbf{I} + \mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\mathbf{A}[\mathbf{A}'\mathbf{X}'(\mathbf{X}\tilde{\boldsymbol{\Delta}}\mathbf{X}' + \mathbf{C})\mathbf{X}\mathbf{A}]^{-1}\mathbf{A}'(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\tilde{\boldsymbol{\Delta}}^{-1} - \tilde{\boldsymbol{\Delta}}^{-1}) \\ &\quad \cdot [\mathbf{X}'\mathbf{C}^{-1}\mathbf{X} + \tilde{\boldsymbol{\Delta}}^{-1}]^{-1}\} \\ &= [\mathbf{X}'\mathbf{C}^{-1}\mathbf{X} + \tilde{\boldsymbol{\Delta}}^{-1}]^{-1} + \tilde{\boldsymbol{\Delta}}\mathbf{X}'(\mathbf{X}\tilde{\boldsymbol{\Delta}}\mathbf{X}' + \mathbf{C})^{-1}\mathbf{X}\mathbf{A}[\mathbf{A}'\mathbf{X}'(\mathbf{X}\tilde{\boldsymbol{\Delta}}\mathbf{X}' + \mathbf{C})^{-1}\mathbf{X}\mathbf{A}]^{-1} \\ &\quad \cdot \mathbf{A}'\mathbf{X}'(\mathbf{X}\tilde{\boldsymbol{\Delta}}\mathbf{X}' + \mathbf{C})^{-1}\mathbf{X}\tilde{\boldsymbol{\Delta}}. \end{aligned} \quad (105)$$

Substituting (105) into (103) we have

$$\begin{aligned}
\tilde{\alpha} &= [\mathbf{X}'\mathbf{C}^{-1}\mathbf{X} + \tilde{\mathbf{\Delta}}^{-1}]^{-1}\mathbf{X}'\mathbf{C}^{-1}\mathbf{y} \\
&\quad -(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X} + \tilde{\mathbf{\Delta}}^{-1})^{-1}(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X} + \tilde{\mathbf{\Delta}}^{-1} - \tilde{\mathbf{\Delta}}^{-1})\mathbf{A}(\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{y} \\
&\quad +\tilde{\mathbf{\Delta}}\mathbf{X}'(\mathbf{X}\tilde{\mathbf{\Delta}}\mathbf{X}' + \mathbf{C})^{-1}\mathbf{X}\mathbf{A}[\mathbf{A}'\mathbf{X}'(\mathbf{X}\tilde{\mathbf{\Delta}}\mathbf{X}' + \mathbf{C})^{-1}\mathbf{X}\mathbf{A}]^{-1}\mathbf{A}'\mathbf{X}'[\mathbf{C}^{-1} - (\mathbf{X}\tilde{\mathbf{\Delta}}\mathbf{X}' + \mathbf{C})^{-1}]\mathbf{y} \\
&\quad -\tilde{\mathbf{\Delta}}\mathbf{X}'(\mathbf{X}\tilde{\mathbf{\Delta}}\mathbf{X}' + \mathbf{C})^{-1}\mathbf{X}\mathbf{A}[\mathbf{A}'\mathbf{X}'(\mathbf{X}\tilde{\mathbf{\Delta}}\mathbf{X}' + \mathbf{C})^{-1}\mathbf{X}\mathbf{A}]^{-1} \\
&\quad \cdot[\mathbf{I} - \mathbf{A}'\mathbf{X}(\mathbf{X}\tilde{\mathbf{\Delta}}\mathbf{X}' + \mathbf{C})^{-1}\mathbf{X}\mathbf{A}](\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{y} \\
&= (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X} + \tilde{\mathbf{\Delta}}^{-1})^{-1}\mathbf{X}'\mathbf{C}^{-1}\mathbf{y} - \mathbf{A}(\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{y} \\
&\quad +(\mathbf{X}'\mathbf{C}^{-1} + \tilde{\mathbf{\Delta}}^{-1})^{-1}\tilde{\mathbf{\Delta}}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{y} \\
&\quad -\tilde{\mathbf{\Delta}}\mathbf{X}'(\mathbf{X}\tilde{\mathbf{\Delta}}\mathbf{X}' + \mathbf{C})^{-1}\mathbf{X}\mathbf{A}[\mathbf{A}'\mathbf{X}'(\mathbf{X}\tilde{\mathbf{\Delta}}\mathbf{X}' + \mathbf{C})^{-1}\mathbf{X}\mathbf{A}]^{-1}\mathbf{A}'\mathbf{X}'(\mathbf{X}\tilde{\mathbf{\Delta}}\mathbf{X}' + \mathbf{C})^{-1}\mathbf{y} \\
&\quad +\tilde{\mathbf{\Delta}}\mathbf{X}'(\mathbf{X}\tilde{\mathbf{\Delta}}\mathbf{X}' + \mathbf{C})^{-1}\mathbf{X}\mathbf{A}(\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}'\mathbf{C}^{-1}\mathbf{y} \\
&= (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X} + \tilde{\mathbf{\Delta}}^{-1})^{-1}\mathbf{X}'\mathbf{C}^{-1}\mathbf{y} - \tilde{\mathbf{\Delta}}\mathbf{X}'(\mathbf{X}\tilde{\mathbf{\Delta}}\mathbf{X}' + \mathbf{C})^{-1}\mathbf{X}\mathbf{A}\hat{\tilde{\beta}}. \tag{106}
\end{aligned}$$

References

- [1] Ahn, S.G., Y.H. Lee and P. Schmidt, (2001), GMM Estimation of Linear Panel Data Models with Time-varying Individual Effects, *Journal of Econometrics*, 102, 219-255.
- [2] Amemiya, T. (1978), “A Note on a Random Coefficients Model”, *International Economic Review*, 19, 793-796.
- [3] Anderson, T.W. (1985), “An Introduction to Multivariate Analysis”, 2nd edition, Wiley, New York.
- [4] Anderson, T.W. and C. Hsiao (1981), “Estimation of Dynamic Models with Error Components”, *Journal of the American Statistical Society*, 76, 598-606.
- [5] Anderson, T.W. and C. Hsiao (1982), “Formulation and Estimation of Dynamic Models Using Panel Data”, *Journal of Econometrics*, 18, 47-82.
- [6] Anselin, L. (2001), “Spatial Econometrics”, in B. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Blackwell, Oxford.
- [7] Bai, J. and S. Ng (2002), “A Panic Attack on Unit Roots and Cointegration”, Department of Economics, Boston College, Unpublished Manuscript.
- [8] Balestra, P. (1991), “Introduction to Linear Models for Panel Data”, incomplete.
- [9] Balestra, P. and S. Negassi (1990), “A Random Coefficient Simultaneous Equation System with an Application to Direct Foreign Investment by French Firms”, working paper 90-05, Département d'économétrie, Université de Genève.
- [10] Boskin, M.J. and L.J. Lau (1990), “Post-War Economic Growth in the Group-of-Five Countries: A New Analysis”, CEPR No. 217, Stanford University.
- [11] Conley, T.G. and G. Topa (2002), “Socio-economic Distance and Spatial Patterns in Unemployment”, *Journal of Applied Econometrics*, 17, 303-327.

- [12] Cooley, T.F. and E.C. Prescott (1976) “Estimation in the Presence of Stochastic Parameter Variation”, *Econometrica*, 44, 167-184.
- [13] de Finetti, B. (1964), “Foresight: Its Logical Laws. Its Subjective Sources”, in *Studies in Subjective Probability*, ed. by J.E. Kyburg, Jr., and H.E. Smokler, New York, Wiley, 93-158.
- [14] Friedman, M. (1953), *Essays in Positive Economics*, Chicago: University of Chicago Press.
- [15] Geisser, S. (1980), “A Predictivistic Primer”, in *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*, Amsterdam: North Holland, 363-382.
- [16] Hausman, J.A. (1978), “Specification Tests in Econometrics”, *Econometrica*, 46, 1251-1271.
- [17] Hendricks, W., R. Koenker and D.J. Poirier (1979), “Residential Demand for Electricity: An Econometric Approach”, *Journal of Econometrics*, 9, 33-57.
- [18] Hildreth, C. and J.P. Houck (1968), “Some Estimators for a Linear Model with Random Coefficients”, *Journal of the American Statistical Association*, 63, 584-595.
- [19] Holtz-Eakin, D, W.K. Newey and H.Rosen (1988), “Estimating Vector Autoregressions with Panel Data”, *Econometrica*, 56, 1371-1395.
- [20] Hsiao, C. (1974), “Statistical Inference for a Model with Both Random Cross-Sectional and Time Effects”, *International Economic Review*, 15, 12-30.
- [21] Hsiao, C. (1975), “Some Estimation Methods for a Random Coefficients Model”, *Econometrica*, 43, 305-325.
- [22] Hsiao, C. (1986), *Analysis of Panel Data*, Econometric Society monographs No. 11, New York: Cambridge University Press.
- [23] Hsiao, C. (1990), “A Mixed Fixed and Random Coefficients Framework for Pooling Cross-section and Time Series Data”, paper presented at the Third

Conference on Telecommunication Demand Analysis with Dynamic Regulation, Hilton, Head, S. Carolina.

- [24] Hsiao, C. (2003), *Analysis of Panel Data*, Economic Society monographs no. 34, 2nd edition, New York: Cambridge University Press.
- [25] Hsiao, C. and B.H. Sun (2000), “To Pool or Not to Pool Panel Data”, in *Panel Data Econometrics: Future Directions*, Papers in Honour of Professor Pietro Balestra, ed. by J. Krishnakumar and E. Ronchetti, Amsterdam: North Holland, 181-198.
- [26] Hsiao, C., T.W. Appelbe, and C.R. Dineen (1992), “A General Framework for Panel Data Models - With an Application to Canadian Customer-Dialed Long Distance Telephone Service”, *Journal of Econometrics* (forthcoming).
- [27] Hsiao, C., D.C. Mountain, K.Y. Tsui and M.W. Luke Chan (1989), “Modeling Ontario Regional Electricity System Demand Using a Mixed Fixed and Random Coefficients Approach”, *Regional Science and Urban Economics* 19, 567-587.
- [28] Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu (1999), “Bayes Estimation of Short-Run Coefficients in Dynamic Panel Data Models”, in *Analysis of Panels and Limited Dependent Variables Models*, by C. Hsiao, L.F. Lee, K. Lahiri and M.H. Pesaran, Cambridge: Cambridge University press, 268-296.
- [29] Hurwicz, L. (1950), “Least Squares Bias in Time Series”, in T.C. Koopman, ed., *Statistical Inference in Dynamic Economic Models*, New York: Wiley, 365-383.
- [30] Jeffreys, H. (1961), *Theory of Probability*, 3rd ed., Oxford: Clarendon Press.
- [31] Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl and T.C. Lee (1985), *The Theory and Practice of Econometrics*, 2nd ed., New York: Wiley.
- [32] Kelejian, H.H. (1974), “Random Parameters in Simultaneous Equation Framework: Identification and Estimation”, *Econometrica*, 42, 517-527.

- [33] Kiviet, J.F. and G.D.A. Phillips (1993), “Alternative Bias Approximation with Lagged Dependent Variables”, *Econometric Theory*, 9, 62-80.
- [34] Klein, L.R. (1988), “The Statistical Approach to Economics”, *Journal of Econometrics*, 37, 7-26.
- [35] Lee, K.C., and Pesaran, M.H. (1993), “The Role of Sectoral Interactions in Wage Determination in the UK Economy”, *The Economic Journal*, 103, 21-55.
- [36] Lindley, D.V. (1961) “The Use of Prior Probability Distributions in Statistical Inference and Decision”, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman, Berkeley: University of California Press, 453-468.
- [37] Lindley, D.V. and A.F.M. Smith (1972), “Bayes Estimates for the Linear Model”, *Journal of the Royal Statistical Society*, B. 34, 1-41.
- [38] Liu, L.M. and G.C. Tiao (1980), “Random Coefficient First-Order Autoregressive Models”, *Journal of Econometrics*, 13, 305-325.
- [39] Min, C.K. and A. Zellner (1993), “Bayesian and Non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rate”, *Journal of Econometrics*, 56, 89-118.
- [40] Moon, H.R. and B. Perron, (2003), “Testing for a Unit Root in Panels with Dynamic Factors”, *Journal of Econometrics* (forthcoming).
- [41] Moon, H.R., B. Perron and P.C.B. Phillips (2003), “Power Comparisons of Panel Unit Root Tests under Incidental Trends”, Department of Economics, *University of Southern California*, unpublished manuscript.
- [42] Pagan, A. (1980), “Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients”, *Journal of Econometrics*, 13, 341-364.
- [43] Pesaran, M.H. (2002), “Estimation and Inference in Large Heterogeneous Panels with Cross Section Dependence”, Cambridge Working Papers in Economics No. 0305 and CESifo Working Paper Series No. 869.

- [44] Pesaran, M.H. (2003), “A Simple Panel Unit Root Test in the Presence of Cross Section Dependence”, Cambridge Working Papers in Economics No. 0346, University of Cambridge.
- [45] Pesaran, M.H. (2004), “General Diagnostic Tests for Cross Section Dependence in Panels”, Unpublished manuscript, Cambridge University.
- [46] Pesaran, M.H. and R. Smith (1995), “Estimation of Long-Run Relationships from Dynamic Heterogeneous Panels”, *Journal of Econometrics*, 68, 79-114.
- [47] Pesaran, M.H., R. Smith and K.S. Im (1996), “Dynamic Linear Models for Heterogeneous Panels”, chapter 8 in Måtyàs, L. and P. Sevestre, eds (1996) *The Econometrics of Panel Data: A Handbook of Theory with Applications*, 2nd revised edition, Dordrecht: Kluwer Academic Publications.
- [48] Pesaran, M.H., Y. Shin and R.P. Smith, (1999), “Pooled Mean Group Estimation of Dynamic Heterogeneous Panels”, *Journal of the American Statistical Association*, 94, 621-634.
- [49] Pesaran, M.H., T. Schuermann and S.M. Weiner (2004), “Modeling Regional Interdependencies using a Global Error-Correcting Macroeconomic Model”, *Journal of Business Economics and Statistics* (with Discussions), forthcoming.
- [50] Phillips, P.C.B. and D. Sul (2002), “Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence”, Cowles Foundation Discussion Paper No. 1362, Yale University.
- [51] Raj, B. and A. Ullah (1981), *Econometrics: A Varying Coefficients Approach*, Croom Helm: London.
- [52] Rao, C.R. (1970), “Estimation of Heteroscedastic Variances in Linear Models”, *Journal of the American Statistical Association*, 65, 161-172.
- [53] Rao, C.R. (1972), “Estimation of Variance and Covariance Components in Linear Models”, *Journal of the American Statistical Association*, 67, 112-115.

- [54] Rao, C.R. (1973), *“Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley.
- [55] Robertson, D. and J. Symons (2000), “Factor Residuals in SUR Regressions: Estimating Panels Allowing for Cross Sectional Correlation”, Unpublished manuscript, Faculty of Economics and Politics, University of Cambridge.
- [56] Rosenberg, B. (1972), “The Estimation of Stationary Stochastic Regression Parameters Re-examined”, *Journal of the American Statistical Association*, 67, 650-654.
- [57] Rosenberg, B. (1973), “The Analysis of a Cross-Section of Time Series by Stochastically Convergent Parameter Regression”, *Annals of Economic and Social Measurement*, 2, 399-428.
- [58] Schwarz, G. (1978), “Estimating the Dimension of a Model”, *Annals of Statistics*, 6, 461-464.
- [59] Singh, B., A.L. Nagar, N.K. Choudhry, and B. Raj (1976), “On the Estimation of Structural Changes: A Generalisation of the Random Coefficients Regression Model”, *International Economic Review*, 17, 340-361.
- [60] Swamy, P.A.V.B. (1970), “Efficient Inference in a Random Coefficient Regression Model”, *Econometrica*, 38, 311-323.
- [61] Swamy, P.A.V.B. and P.A. Tinsley (1977), “Linear Prediction and Estimation Methods for Regression Models with Stationary Stochastic Coefficients”, *Federal Reserve Board Division of Research and Statistics*, Special Studies Paper No. 78, Washington, D.C.
- [62] Wachter, M.L. (1976), “The Changing Cyclical Responsiveness of Wage Inflation”, *Brookings paper on Economic Activity*, 1976, 115-168.
- [63] Zellner, A. (1962), “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias”, *Journal of the American Statistical Association*, 57, 348-368.

- [64] Zellner, A. (1988), “Bayesian Analysis in Econometrics”, *Journal of Econometrics*, 37, 27-50.