CHAPTER 6

Variable-Coefficient Models

6.1 INTRODUCTION

So far we have confined our discussion to models in which the effects of omitted variables are either individual-specific or time-specific or both. But there are cases in which there are changing economic structures or unobserved different socioeconomic and demographic background factors that imply that the response parameters of the included variables may be varying over time and/or may be different for different cross-sectional units. For instance, in farm production it is likely that variables not included in the specification could also impact the marginal productivity of fertilizer used such as soil characteristics (e.g., slope, soil fertility, water reserve, etc.) or climatic conditions. The same applies to empirical studies of economic growth. The per capita output growth rates are assumed to depend on two sets of variables over a common horizon. One set of variables consists of initial per capita output, savings, and population growth rates, variables that are suggested by the Solow growth model. The second set of variables consists of control variables that correspond to whatever additional determinants of growth a researcher wishes to examine (e.g., Durlauf 2001; Durlauf and Quah 1999). However, there is nothing in growth theory that would lead one to think that the marginal effect of a change in high school enrollment percentages on the per capita growth of the United States should be the same as the effect on a country in sub-Saharan Africa. Had all these factors been taken into account in the specification, a common slope coefficients model may seem reasonable. However, these variables could be unavailable or could be difficult to observe with precision. Moreover, a model is not a mirror; it is a simplification of the real world to capture the relationships among the essential variables. As a matter of fact, any parsimonious regression will necessarily leave out many factors that would, from the perspective of economic theory, be likely to affect the parameters of the included variables (e.g., Canova 1999; Durlauf and Johnson 1995). In these situations, varying parameter models appear to be more capable of capturing the unobserved heterogeneity than a model with only individual-specific and/or time-specific effects (variable-intercept models).

In Chapter 2 we reported a study (Kuh 1963) on investment expenditures of 60 small and middle-sized firms in capital-goods-producing industries

from 1935 to 1955, excluding the war years (1942–45). In a majority of the cases Kuh investigated, the common intercept and common slope coefficients for all firms, as well as the variable-intercept common-slope hypotheses, were rejected (Tables 2.3 and 2.4). Similar results were found by Swamy (1970), who used the annual data of 11 U.S. corporations from 1935 to 1954 to fit the Grunfeld (1958) investment functions. His preliminary test of variable-intercept–common slope coefficients against the variable-intercept and variable slope coefficients for the value of a firm's outstanding shares at the beginning of the year and its beginning-of-year capital stock yielded an F value of 14.4521. That is well above the 5 percent value of an F distribution with 27 and 187 degrees of freedom.[1]

When an investigator is interested mainly in the fundamental relationship between the outcome variable and a set of primary conditional variables, either for ease of analysis or because of the unavailability of the secondary conditional variables, it would seem reasonable to allow variations in parameters across cross-sectional units and/or over time as a means to take account of the interindividual and/or interperiod heterogeneity. A single-equation model in its most general form can be written as

$$y_{it} = \sum_{k=1}^{K} \beta_{kit} x_{kit} + u_{it}, \quad i = 1, \ldots, N, \quad t = 1. \ldots, T, \tag{6.1.1}$$

where, in contrast to previous chapters, we no longer treat the intercept differently than other explanatory variables and let $x_{1it} = 1$. However, if all the coefficients are treated as fixed and different for different cross-sectional units in different time periods, there are NKT parameters with only NT observations. Obviously, there is no way we can obtain any meaningful estimates of β_{kit}. We are thus led to search for an approach that allows the coefficients of interest to differ, but provides some method of modeling the cross-sectional units as a group rather than individually.

One possibility would be to introduce dummy variables into the model that would indicate differences in the coefficients across individual units and/or over time, that is, to develop an approach similar to the least-squares dummy variable approach. In the case in which only cross-sectional differences are present, this approach is equivalent to postulating a separate regression for each cross-sectional unit[2]

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}_i + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{6.1.2}$$

where $\boldsymbol{\beta}_i$ and $\mathbf{x}_{it}$ are $K \times 1$ vectors of parameters and explanatory variables.

[1] See Mehta, Narasimham, and Swamy (1978) for another example that using error-components formulation to account for heterogeneity does not always yield economically meaningful results.

[2] Alternatively, we can postulate a separate regression for each time period; so $y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}_t + u_{it}$.

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Alternatively, each regression coefficient can be viewed as a random variable with a probability distribution (e.g., Hurwicz 1950; Klein 1953; Theil and Mennes 1959; Zellner 1966). The random-coefficients 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. Depending on the type of assumption about the parameter variation, it can be further classified into one of two categories: stationary and nonstationary random-coefficient models.

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

$$\begin{aligned} \boldsymbol{\beta}_{it} &= \bar{\boldsymbol{\beta}} + \boldsymbol{\xi}_{it}, \quad i = 1, \ldots, N, \\ &\qquad\qquad t = 1, \ldots, T, \end{aligned} \tag{6.1.3}$$

where $\bar{\boldsymbol{\beta}}$ is a $K \times 1$ vector of constants, and $\boldsymbol{\xi}_{it}$ is a $K \times 1$ vector of stationary random variables with 0 means and constant variance–covariances. For this type of model we are interested in (1) estimating the mean coefficient vector $\bar{\boldsymbol{\beta}}$, (2) predicting each individual component $\boldsymbol{\xi}_{it}$, (3) estimating the dispersion of the individual-parameter vector, and (4) testing the hypothesis that the variances of $\boldsymbol{\xi}_{it}$ are 0.

The nonstationary random-coefficient models do not regard the coefficient vector as having constant mean or variance. 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. In this case we are interested in (1) estimating the parameters characterizing the time-evolving process, (2) estimating the initial value and the history of parameter realizations, (3) predicting the future evolutions, and (4) testing the hypothesis of random variation.

Because of the computational complexities, variable-coefficient models have not gained as wide acceptance in empirical work as has the variable-intercept model. However, that does not mean that there is less need for taking account of parameter heterogeneity in pooling the data. In this chapter we survey some of the popular single-equation varying coefficients models. We first discuss models in which the variations of coefficients are independent of the variations of exogenous explanatory variables. Single-equation models with coefficients varying over individuals are discussed in Section 6.2. In Section 6.3 we discuss models with coefficients varying over individuals and time. Section 6.4 concerns models with time-evolving coefficients. Models with coefficients that are functions of other exogenous variables are discussed in Section 6.5. Section 6.6 proposes a mixed fixed and random coefficient model as a unifying framework to various approaches of controlling unobserved heterogeneity. Section 6.7 discusses issues of dynamic models. Section 6.8 provides two examples that use random coefficients model to pool heterogeneous individuals. General models in which the variation of coefficients are correlated with the explanatory

variables are discussed in Section 6.9. For random coefficients models with heteroscedasticity, see Bresson et al. (2011); with cross-correlated residuals (e.g., Bresson and Hsiao 2011); simultaneous-equations models with random coefficients, see Chow (1983), Kelejian (1977), and Raj and Ullah (1981). Further discussion of this subject can also be found in Amemiya (1983), Chow (1983), Hsiao and Pesaran (2008), Judge et al. (1980), and Raj and Ullah (1981).

6.2 COEFFICIENTS THAT VARY OVER CROSS-SECTIONAL UNITS

When regression coefficients are viewed as invariant over time, but varying from one unit to another, we can write the model as

$$\begin{aligned} y_{it} &= \sum_{k=1}^{K} \beta_{ki} x_{kit} + u_{it} \\ &= \sum_{k=1}^{K} (\bar{\beta}_k + \alpha_{ki}) x_{kit} + u_{it}, \quad \begin{matrix} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{matrix} \end{aligned} \tag{6.2.1}$$

where $\bar{\boldsymbol{\beta}} = (\bar{\beta}_1, \ldots, \bar{\beta}_K)'$ can be viewed as the common-mean-coefficient vector and $\boldsymbol{\alpha}_i = (\alpha_{1i}, \ldots, \alpha_{Ki})'$ as the individual deviation from the common mean $\bar{\boldsymbol{\beta}}$. If individual observations are heterogeneous or the performance of individual units from the data base is of interest, then $\boldsymbol{\alpha}_i$ are treated as fixed constants. If conditional on x_{kit}, individual units can be viewed as random draws from a common population or the population characteristics are of interest, then α_{ki} are generally treated as random variables having 0 means and constant variances and covariances.

6.2.1 Fixed-Coefficient Model

6.2.1.1 Complete Heterogeneity

When $\boldsymbol{\beta}_i$ are treated as fixed and different constants, we can stack the NT observations in the form of the Zellner (1962) seemingly unrelated regression model

$$\begin{aligned} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} &= \begin{bmatrix} X_1 & & & \mathbf{0} \\ & X_2 & & \\ & & \ddots & \\ \mathbf{0} & & & X_N \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_N \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_N \end{bmatrix}, \\ &= \tilde{X}\boldsymbol{\beta} + \mathbf{u} \end{aligned} \tag{6.2.2}$$

where $\mathbf{y}_i$ and $\mathbf{u}_i$ are $T \times 1$ vectors of $(y_{i1}, \ldots, y_{iT})'$ and $(u_{i1}, \ldots, u_{iT})'$; X_i is the $T \times K$ matrix of the time-series observations of the ith individual's explanatory variables with the tth row equal to $\mathbf{x}_{it}'$; $\tilde{X}$ is $NT \times NK$ block

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

diagonal matrix with the ith block being X_i; and $\boldsymbol{\beta}$ is an $NK \times 1$ vector, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_N')'$. If the covariances between different cross-sectional units are not 0, $E\mathbf{u}_i\mathbf{u}_j' \neq \mathbf{0}$, the GLS estimator of $(\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_N')$ is more efficient than the single-equation estimator of $\boldsymbol{\beta}_i$ for each cross-sectional unit. If X_i are identical for all i or $E\mathbf{u}_i\mathbf{u}_i' = \sigma_i^2 I$ and $E\mathbf{u}_i\mathbf{u}_j' = \mathbf{0}$ for $i \neq j$, the generalized least-squares (GLS) estimator for $(\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_N')$ is the same as applying the least squares separately to the time-series observations of each cross-sectional unit.

6.2.1.2 Group Heterogeneity

When N is large, it is neither feasible nor desirable to let $\boldsymbol{\beta}_i$ be different for different i. An alternative to individual heterogeneity is to assume group heterogeneity in place of complete heterogeneity. In other words, the population is assumed to be composed of G heterogeneous groups. Individuals belonging to a particular group all have the same response function (e.g., Lin and Ng 2012; Su, Shi, and Phillips 2013),

$$y_{it,g} = \mathbf{x}_{it}'\boldsymbol{\beta}_g + u_{it,g}, \quad \text{for } i \in \text{group } g. \tag{6.2.3}$$

If the grouping is known from some external consideration (e.g., Bester and Hansen 2012), estimation of (6.2.3) can proceed following the Zellner (1962) seemingly unrelated regression framework. However, if such external information is not available, two issues arise: (1) how to determine the number of groups, G; and (2) how to identify the group to which an individual belongs. Following the idea of Lasso (Least Absolute Shrinkage and Selection Operator; Tibshirani (1996)) under the assumption that the number of groups, G, is known, Su, Shi, and Phillips (SSP) (2013) suggest a modified penalized least-squares approach,

$$\text{Min } Q^G = Q + \frac{a}{N}\sum_{i=1}^{N}\prod_{g=1}^{G} \parallel \boldsymbol{\beta}_i - \boldsymbol{\beta}_g \parallel, \tag{6.2.4}$$

to simultaneously classify individuals into groups and estimate $\boldsymbol{\beta}_g$, where $\| \cdot \|$ denotes the Frobenius norm, $\| A \| = [tr\ AA']^{1/2}$,

$$Q = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \mathbf{x}_{it}'\boldsymbol{\beta}_i)^2.$$

and a is a tuning constant. SSP show that minimizing (6.2.4) achieves classification of individuals into groups and consistent estimation of $\boldsymbol{\beta}_g$ in a single step when N and T are large. SSP also propose to select the number of groups, G, by minimizing

$$\log \hat{\sigma}_G^2 + cGK, \tag{6.2.5}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

where

$$\hat{\sigma}_G^2 = \frac{1}{NT}\sum_{g=1}^{G}\sum_{i\epsilon g}\sum_{t=1}^{T}(y_{it} - \mathbf{x}_{it}'\hat{\boldsymbol{\beta}}_{g,G})^2$$

is the estimated average residual sum of squares based on G-group estimates of (6.2.3), $\hat{\boldsymbol{\beta}}_{g,G}$, and c is a turning constant.

6.2.2 Random-Coefficient Model

6.2.2.1 The Model

When $\boldsymbol{\beta}_i = \bar{\boldsymbol{\beta}} + \boldsymbol{\alpha}_i$ are treated as random, with common mean $\bar{\boldsymbol{\beta}}$, Swamy (1970) assumed that[3]

$$\begin{aligned} E\boldsymbol{\alpha}_i &= \mathbf{0}, \\ \underset{K\times K}{E\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j'} &= \begin{cases} \Delta & \text{if } i = j, \\ \mathbf{0} & \text{if } i \neq j, \end{cases} \\ E\mathbf{x}_{it}\boldsymbol{\alpha}_j' &= \mathbf{0}, \quad E\boldsymbol{\alpha}_i\mathbf{u}_j' = \mathbf{0}, \\ E\mathbf{u}_i\mathbf{u}_j' &= \begin{cases} \sigma_i^2 I_T & \text{if } i = j, \\ \mathbf{0} & \text{if } i \neq j. \end{cases} \end{aligned} \tag{6.2.6}$$

Stacking all NT observations, we have

$$\mathbf{y} = X\bar{\boldsymbol{\beta}} + \tilde{X}\boldsymbol{\alpha} + \mathbf{u}, \tag{6.2.7}$$

where

$$\underset{NT\times 1}{\mathbf{y}} = (\mathbf{y}_1', \ldots, \mathbf{y}_N')',$$

$$\underset{NT\times K}{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}, \quad \underset{NT\times NK}{\tilde{X}} = \begin{bmatrix} X_1 & & & \mathbf{0} \\ & X_2 & & \\ & & \ddots & \\ \mathbf{0} & & & X_N \end{bmatrix} = \operatorname{diag}(X_1, \ldots, X_N),$$

$\mathbf{u} = (\mathbf{u}_1', \ldots, \mathbf{u}_N')'$, and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_N')'$. The covariance matrix for the composite disturbance term $\tilde{X}\boldsymbol{\alpha} + \mathbf{u}$ is block-diagonal, with the ith diagonal block given by

$$\Phi_i = X_i\Delta X_i' + \sigma_i^2 I_T. \tag{6.2.8}$$

[3] See Chamberlain (1992) for an extension of the Mundlak–Chamberlain approach of conditioning the individual effects on the conditioning variables to models with individual-specific slopes that may be correlated with conditioning variables. An instrumental variable estimator is proposed within a finite dimensional, method of moments framework. Also, see section 6.9.

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

6.2.2.2 *Estimation*

Under Swamy's (1970) assumption, the simple regression of $\mathbf{y}$ on X will yield an unbiased and consistent estimator of $\bar{\boldsymbol{\beta}}$ if $(1/NT)\, X'X$ converges to a nonsingular constant matrix. But the estimator is inefficient, and the usual least-squares formula for computing the variance–covariance matrix of the estimator is incorrect, often leading to misleading statistical inferences. Moreover, when the pattern of parameter variation is of interest in its own right, an estimator ignoring parameter variation is incapable of shedding light on this aspect of the economic process.

The best linear unbiased estimator of $\bar{\boldsymbol{\beta}}$ for (6.2.7) is the GLS estimator[4]

$$\begin{aligned}\hat{\bar{\boldsymbol{\beta}}}_{GLS} &= \left(\sum_{i=1}^{N} X_i' \Phi_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i' \Phi_i^{-1} \mathbf{y}_i\right) \\ &= \sum_{i=1}^{N} W_i \hat{\boldsymbol{\beta}}_i,\end{aligned} \tag{6.2.9}$$

where

$$W_i = \left\{\sum_{i=1}^{N} [\Delta + \sigma_i^2 (X_i' X_i)^{-1}]^{-1}\right\}^{-1} [\Delta + \sigma_i^2 (X_i' X_i)^{-1}]^{-1},$$

and

$$\hat{\boldsymbol{\beta}}_i = (X_i' X_i)^{-1} X_i' \mathbf{y}_i.$$

The last expression of (6.2.9) shows that the GLS estimator is a matrix-weighted average of the least-squares estimator for each cross-sectional unit, 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 simple least-squares estimator.

[4] Repeatedly using the formula that $(A + BDB')^{-1} = A^{-1} - A^{-1}B(B'A^{-1}B + D^{-1})^{-1}B'A^{-1}$ (Rao 1973, their Chapter 1), we have

$$\begin{aligned} X_i' \Phi_i^{-1} X_i &= X_i' [\sigma_i^2 I + X_i \Delta X_i']^{-1} X_i \\ &= X_i' \left\{\frac{1}{\sigma_i^2} I_T - \frac{1}{\sigma_i^2} X_i [X_i' X_i + \sigma_i^2 \Delta^{-1}]^{-1} X_i'\right\} X_i \\ &= \frac{1}{\sigma_i^2}\left[X_i' X_i - X_i' X_i \left\{(X_i' X_i)^{-1}\right.\right. \\ &\quad \left.\left. - (X_i' X_i)^{-1}\left[(X_i' X_i)^{-1} + \frac{1}{\sigma_i^2}\Delta\right]^{-1} (X_i' X_i)^{-1}\right\} X_i' X_i\right] \\ &= [\Delta + \sigma_i^2 (X_i' X_i)^{-1}]^{-1}. \end{aligned}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

The covariance matrix for the GLS estimator is

$$\begin{aligned}\text{Var}\,(\hat{\bar{\boldsymbol{\beta}}}_{GLS}) &= \left(\sum_{i=1}^{N} X_i'\Phi_i^{-1}X_i\right)^{-1} \\ &= \left\{\sum_{i=1}^{N}[\Delta + \sigma_i^2(X_i'X_i)^{-1}]^{-1}\right\}^{-1}.\end{aligned} \tag{6.2.10}$$

Swamy (1970) proposed using the least-squares estimators $\hat{\boldsymbol{\beta}}_i = (X_i'X_i)^{-1}X_i'\mathbf{y}_i$ and their residuals $\hat{\mathbf{u}}_i = \mathbf{y}_i - X_i\hat{\boldsymbol{\beta}}_i$ to obtain unbiased estimators of σ_i^2 and Δ,[5]

$$\begin{aligned}\hat{\sigma}_i^2 &= \frac{\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i}{T-K} \\ &= \frac{1}{T-K}\mathbf{y}_i'[I - X_i(X_i'X_i)^{-1}X_i']\mathbf{y}_i,\end{aligned} \tag{6.2.11}$$

$$\begin{aligned}\hat{\Delta} = \frac{1}{N-1}\sum_{i=1}^{N}&\left(\hat{\boldsymbol{\beta}}_i - N^{-1}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i\right) \\ &\cdot\left(\hat{\boldsymbol{\beta}}_i - N^{-1}\sum_{i=1}^{N}\hat{\beta}_i\right)' - \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_i^2(X_i'X_i)^{-1}.\end{aligned} \tag{6.2.12}$$

Again, just as in the error-component model, the estimator (6.2.12) is not necessarily nonnegative definite. In this situation, Swamy (see also Judge et al. 1980) has suggested replacing (6.2.12) by

$$\hat{\Delta} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_i - N^{-1}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i\right)\left(\hat{\boldsymbol{\beta}}_i - N^{-1}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i\right)'. \tag{6.2.13}$$

This estimator, although not unbiased, is non-negative definite and is consistent when both N and T tend to infinity. Alternatively, we can use the Bayes mode estimator suggested by Lindley and Smith (1972) and Smith (1973),

$$\Delta^* = \{R + (N-1)\hat{\Delta}\}/(N + \rho - K - 2), \tag{6.2.14}$$

where R and ρ are prior parameters, assuming that Δ^{-1} has a Wishart distribution with ρ degrees of freedom and matrix R. For instance, we may let $R = \hat{\Delta}$ and $\rho = 2$ as in Hsiao, Pesaran, and Tahmiscioglu (1999).

Swamy (1970) proved that substituting $\hat{\sigma}_i^2$ and $\hat{\Delta}$ for σ_i^2 and Δ in (6.2.9) yields an asymptotically normal and efficient estimator of $\bar{\boldsymbol{\beta}}$. The speed of convergence of the GLS estimator is $N^{1/2}$. This can be seen by noting that the

[5] Equation (6.2.9) follows from the relation that $\hat{\boldsymbol{\beta}}_i = \boldsymbol{\beta}_i + (X_i'X_i)^{-1}X'\mathbf{u}_i$ and $E(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})' = \Delta + \sigma_i^2(X_i'X_i)^{-1}$.

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

inverse of the covariance matrix for the GLS estimator [equation (6.2.10)] is[6]

$$\begin{aligned}\operatorname{Var}(\hat{\bar{\boldsymbol{\beta}}}_{\mathrm{GLS}})^{-1} &= N\Delta^{-1} - \Delta^{-1}\left[\sum_{i=1}^{N}\left(\Delta^{-1} + \frac{1}{\sigma_i^2}X_i'X_i\right)^{-1}\right]\Delta^{-1} \\ &= O(N) - O(N/T).\end{aligned} \tag{6.2.15}$$

Swamy (1970) used the model (6.2.6) and (6.2.7) to reestimate the Grunfeld investment function with the annual data of 11 U.S. corporations. His GLS estimates of the common-mean coefficients of the firms' beginning-of-year value of outstanding shares and capital stock are 0.0843 and 0.1961, with asymptotic standard errors 0.014 and 0.0412, respectively. The estimated dispersion measure of these coefficients is

$$\hat{\Delta} = \begin{bmatrix} 0.0011 & -0.0002 \\ & 0.0187 \end{bmatrix}. \tag{6.2.16}$$

Zellner (1966) has shown that when each $\boldsymbol{\beta}_i$ can be viewed as a random variable with a constant mean, and $\boldsymbol{\beta}_i$ and $\mathbf{x}_i$ are uncorrelated, thereby satisfying Swamy's (1970) assumption, the model will not possess an aggregation bias. In this sense, Swamy's estimate can also be interpreted as an average relationship indicating that in general the value of a firm's outstanding shares is an important variable explaining the investment.

6.2.2.3 Predicting Individual Coefficients

Sometimes one may wish to predict the individual component $\boldsymbol{\beta}_i$, because it provides information on the behavior of each individual and also because it provides a basis for predicting future values of the dependent variable for a given individual. Swamy (1970, 1971) has shown that the best linear unbiased predictor, conditional on given $\boldsymbol{\beta}_i$, is the least-squares estimator $\hat{\boldsymbol{\beta}}_i$. However, if the sampling properties of the class of predictors are considered in terms of repeated sampling over both time and individuals, Lee and Griffiths (1979) (see also Lindley and Smith 1972 and Section 6.6) have suggested predicting $\boldsymbol{\beta}_i$ by

$$\hat{\boldsymbol{\beta}}_i^* = \hat{\bar{\boldsymbol{\beta}}}_{\mathrm{GLS}} + \Delta X_i'(X_i\Delta X_i' + \sigma_i^2 I_T)^{-1}(\mathbf{y}_i - X_i\hat{\bar{\boldsymbol{\beta}}}_{\mathrm{GLS}}). \tag{6.2.17}$$

This predictor is the best linear unbiased estimator in the sense that $E(\hat{\boldsymbol{\beta}}_i^* - \boldsymbol{\beta}_i) = \mathbf{0}$, where the expectation is an unconditional one.

6.2.2.4 Testing for Coefficient Variation

An important question in empirical investigation is whether or not the regression coefficients are indeed varying across cross-sectional units. Because the

[6] We use the notation $O(N)$ to denote that the sequence $N^{-1}a_N$ is bounded (Theil 1971, p. 358).

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

effect of introducing random-coefficient variation is to give the dependent variable a different variance at each observation, models with this feature can be transformed into a particular heteroscedastic formulation, and likelihood-ratio tests can be used to detect departure from the constant-parameter assumption. However, computation of the likelihood-ratio test statistic can be complicated. To avoid the iterative calculations necessary to obtain maximum-likelihood estimates of the parameters in the full model, Breusch and Pagan (1979) have proposed a Lagrangian multiplier test for heteroscedasticity. Their test has the same asymptotic properties as the likelihood-ratio test in standard situations, but it is computationally much simpler. It can be computed simply by repeatedly applying least-square regressions.

Dividing the individual-mean-over-time equation by σ_i^{-1}, we have

$$\frac{1}{\sigma_i}\bar{y}_i = \frac{1}{\sigma_i}\bar{\mathbf{x}}_i'\bar{\boldsymbol{\beta}} + \omega_i, \quad i = 1., \ldots, N, \tag{6.2.18}$$

where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it}$, $\bar{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}$,

$$\omega_i = \frac{1}{\sigma_i}\bar{\mathbf{x}}_i'\boldsymbol{\alpha}_i + \frac{1}{\sigma_i}\bar{u}_i.$$

When the assumption (6.2.6) holds, model (6.2.18) is a model with heteroscedastic variances, $\mathrm{Var}(\omega_i) = (1/T) + (1/\sigma_i^2)\bar{\mathbf{x}}_i'\Delta\bar{\mathbf{x}}_i, i = 1, \ldots, N$. Under the null hypothesis that $\Delta = \mathbf{0}$, (6.2.18) has homoscedastic variances, $\mathrm{Var}(\omega_i) = 1/T, i = 1, \ldots, N$. Thus, we can generalize the Breusch and Pagan (1979) test of heteroscedasticity to test for random-coefficient variation here.

Following the procedures of Rao (1973, pp. 418–19) it can be shown that the transformed Lagrangian-multiplier statistic[7] for testing the null hypothesis leads to computing one-half the predicted sum of squares in a regression of

$$(T\omega_i^2 - 1) = \frac{1}{\sigma_i^2}\left[\sum_{k=1}^{K}\sum_{k'=1}^{K}\bar{x}_{ki}\bar{x}_{k'i}\sigma^2_{\alpha_{kk'}}\right] + \epsilon_i, \quad i = 1, \ldots, N, \tag{6.2.19}$$

where $\sigma^2_{\alpha_{kk'}} = E(\alpha_{ki}\alpha_{k'i})$.[8] Because ω_i and σ_i^2 usually are unknown, we can substitute them by their estimated values $\hat{\omega}_i$ and $\hat{\sigma}_i^2$, where $\hat{\omega}_i$ is the least-squares residual of (6.2.18) and $\hat{\sigma}_i^2$ is given by (6.2.11). When both N and T tend to infinity, the transformed Lagrangian-multiplier statistic has the same

[7] We call this a transformed Lagrangian multiplier test because it is derived by maximizing the log-likelihood function of y_i/σ_i rather than maximizing the log-likelihood function of y_{it}/σ_{it}.

[8] Let

$$(T\hat{\omega}_i^2 - 1) = \frac{1}{\sigma_i^2}\left[\sum_{i=1}^{K}\sum_{i'=1}^{K}\bar{x}_{ki}\bar{x}_{k'i}\hat{\sigma}_{\alpha kk'}\right]$$

be the least-squares predicted value of $(T\hat{\omega}_i^2 - 1)$; then the predicted sum of squares is

$$\sum_{i=1}^{N}(T\hat{\omega}_i^2 - 1)^2.$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

limiting distribution as χ^2 with $[K(K+1)]/2$ degrees of freedom under the null hypothesis of $\Delta = \mathbf{0}$.

The Breusch and Pagan (1979) lagrangian multiplier test can be put into the White (1980) information matrix test framework. Chesher (1984) has shown that the many variants of varying parameters of the same general type of model under consideration can be tested using the statistic

$$D_N(\hat{\boldsymbol{\theta}}_N) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial^2 \log f(y_{it} \mid \mathbf{x}_{it}, \hat{\boldsymbol{\theta}}_N)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\frac{\partial \log f(y_{it} \mid \mathbf{x}_{it}, \hat{\boldsymbol{\theta}}_N)}{\partial\boldsymbol{\theta}}\right]\left[\sum_{t=1}^{T}\frac{\partial \log f(y_{it} \mid \mathbf{x}_{it}; \hat{\boldsymbol{\theta}}_N)}{\partial\boldsymbol{\theta}'}\right], \tag{6.2.20}$$

where $f(y_{it} \mid \mathbf{x}_{it}, \boldsymbol{\theta})$ denotes the conditional density of y_{it} given $\mathbf{x}_{it}$ and $\boldsymbol{\theta}$ under the null of no parameter variation, and $\hat{\boldsymbol{\theta}}_N$ denotes the maximum-likelihood estimator of $\boldsymbol{\theta}$. Under the null, $E\left(\frac{\partial^2 \log f(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right) = -E\left(\frac{\partial \log f(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \cdot \frac{\partial \log f(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right)$. Therefore, elements of $\sqrt{N}D_N(\hat{\boldsymbol{\theta}}_N)$ are asymptotically jointly normal with mean 0 and covariance matrix given by White (1980) and simplified by Chesher (1983) and Lancaster (1984).

Alternatively, because for given i, $\boldsymbol{\alpha}_i$ is fixed, we can test for random variation indirectly by testing whether or not the fixed-coefficient vectors $\boldsymbol{\beta}_i$ are all equal. That is, we form the null hypothesis:

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \ldots = \boldsymbol{\beta}_N = \bar{\boldsymbol{\beta}}.$$

If different cross-sectional units have the same variance, $\sigma_i^2 = \sigma^2, i = 1, \ldots, N$, the conventional analysis-of-covariance (ANCOVA) test for homogeneity discussed in Chapter 2 (F_3) can be applied. If σ_i^2 are assumed different, as postulated by Swamy (1970, 1971), we can apply the modified test statistic

$$F_3^* = \sum_{i=1}^{N}\frac{(\hat{\boldsymbol{\beta}}_i - \hat{\bar{\boldsymbol{\beta}}}^*)' X_i' X_i (\hat{\boldsymbol{\beta}}_i - \hat{\bar{\boldsymbol{\beta}}}^*)}{\hat{\sigma}_i^2}, \tag{6.2.21}$$

where

$$\hat{\bar{\boldsymbol{\beta}}}^* = \left[\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}X_i'X_i\right]^{-1}\left[\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}X_i'\mathbf{y}_i\right].$$

Under H_0, (6.2.21) is asymptotically χ^2 distributed, with $K(N-1)$ degrees of freedom, as T tends to infinity and N is fixed.

Similarly, one can test for slope homogeneity conditional on individual-specific effects. Let $X_i = (\mathbf{e}_T, \tilde{X}_i)$ and $\boldsymbol{\beta}_i' = (\beta_{1i}, \boldsymbol{\beta}_{2i}')$, where $\tilde{X}_i$ denotes the $T \times (K-1)$ time-varying exogenous variables, $\tilde{\mathbf{x}}_{2,it}$ and $\boldsymbol{\beta}_{2i}$ denotes the

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

$(K-1)\times 1$ coefficients of $\mathbf{x}_{2,it}$. Then

$$\tilde{F}_3^* = \sum_{i=1}^{N} \left(\hat{\boldsymbol{\beta}}_{2i} - \hat{\tilde{\boldsymbol{\beta}}}_2^*\right)' \left[\frac{1}{\hat{\sigma}_i^2}\tilde{X}_i' Q \tilde{X}_i\right] \left(\hat{\boldsymbol{\beta}}_{2i} - \hat{\tilde{\boldsymbol{\beta}}}_2^*\right), \tag{6.2.22}$$

where $Q = I_T - \frac{1}{T}\mathbf{e}\mathbf{e}'$,

$$\hat{\boldsymbol{\beta}}_{2i} = (\tilde{X}_i' Q \tilde{X}_i)^{-1}(\tilde{X}_i' Q \mathbf{y}_i), \tag{6.2.23}$$

$$\hat{\tilde{\boldsymbol{\beta}}}_2^* = \left(\sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2}\tilde{X}_i' Q \tilde{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2}\tilde{X}_i' Q y_i\right) \tag{6.2.24}$$

and

$$\hat{\sigma}_i^2 = \frac{1}{T-K}(\mathbf{y}_i - \tilde{\mathbf{X}}_i'\hat{\boldsymbol{\beta}}_{2i})' Q (\mathbf{y}_i - \tilde{X}_i'\hat{\boldsymbol{\beta}}_{2i}). \tag{6.2.25}$$

The statistic $\tilde{F}_3^*$ is asymptotically χ^2 distributed with $(K-1)(N-1)$ degrees of freedom when N is fixed and $T \to \infty$. Pesaran and Yamagata (2008) show that when both N and T go to infinity $\frac{1}{\sqrt{N}}F_3^*$ or $\frac{1}{\sqrt{N}}\tilde{F}_3^*$ is asymptotically normally distributed with mean 0 and variance 1 provided $\frac{\sqrt{N}}{T} \to 0$ as $N \to \infty$. Furthermore, they show that if $\hat{\sigma}_i^2$ ((6.2.25)) is replaced by the estimator

$$\tilde{\sigma}_i^2 = \frac{1}{T-1}(\mathbf{y}_i - \tilde{X}_i\hat{\boldsymbol{\beta}}_{2i})' Q (\mathbf{y}_i - \tilde{X}_i\tilde{\boldsymbol{\beta}}_{2i}), \tag{6.2.26}$$

$\frac{1}{\sqrt{N}}F_3^*$ or $\frac{1}{\sqrt{N}}\tilde{F}_3^*$ possesses better finite sample properties than using $\hat{\sigma}_i^2$ in (6.2.21) or (6.2.22).

6.2.2.5 Fixed or Random Coefficients

The question whether $\boldsymbol{\beta}_i$ should be assumed fixed and different or random and different depends on whether $\boldsymbol{\beta}_i$ can be viewed as from a heterogeneous population or random draws from a common population or whether we are making inferences conditional on the individual characteristics or making unconditional inferences on the population characteristics. If $\boldsymbol{\beta}_i$ are heterogeneous or we are making inferences conditional on the individual characteristics, the fixed-coefficient model should be used. If $\boldsymbol{\beta}_i$ can be viewed as random draws from a common population and inference is on the population characteristics, the random-coefficient model should be used. However, extending his work on the variable-intercept model, Mundlak (1978b) has raised the issue of whether or not the variable coefficients are correlated with the explanatory variables. If they are, the assumptions of the Swamy random-coefficient model are unreasonable, and the GLS estimator of the mean coefficient vector will be biased. To correct this bias, Mundlak (1978b) suggested that the inferences of $f(\mathbf{y}_i \mid X_i, \boldsymbol{\beta})$ be viewed as $\int f(\mathbf{y}_i \mid X_i, \bar{\boldsymbol{\beta}}, \boldsymbol{\alpha}_i) f(\boldsymbol{\alpha}_i \mid X_i)\, d\boldsymbol{\alpha}_i$, where $f(\mathbf{y}_i \mid X_i, \bar{\boldsymbol{\beta}}, \boldsymbol{\alpha}_i)$ denotes the conditional density of y_i given X_i, $\bar{\boldsymbol{\beta}}$ and $\boldsymbol{\alpha}_i$ and

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

$f(\boldsymbol{\alpha}_i \mid X_i)$ denotes the conditional density of $\boldsymbol{\alpha}_i$ given X_i which provides auxiliary equations for the coefficient vector $\boldsymbol{\alpha}_i$ as a function of the ith individual's observed explanatory variables. Because this framework can be viewed as a special case of a random-coefficients model with the coefficients being functions of other explanatory variables, we shall maintain the assumption that the random coefficients are not correlated with the explanatory variables, and we shall discuss estimation of the random coefficients that are functions of other explanatory variables in Section 6.5.

6.2.2.6 An Example

To illustrate the specific issues involved in estimating a behavioral equation using temporal cross-sectional observations when the data do not support the hypothesis that the coefficients are the same for all cross-sectional units, we report a study conducted by Barth, Kraft, and Kraft (1979). They used quarterly observations on output prices, wages, materials prices, inventories, and sales for 17 manufacturing industries for the period 1959 (I) to 1971 (II) to estimate a price equation for the U.S. manufacturing sector. Assuming heteroscedastic disturbance, but common intercept and slope coefficients across industries, and using the two-step Aitken estimator, Barth et al. (1979) obtained

$$\begin{array}{lccccc} \hat{y} = & 0.0005 & + & 0.2853x_2 + & 0.0068x_3 + & 0.0024x_4, \\ & (0.0003) & & (0.0304) & (0.005) & (0.0017) \end{array} \tag{6.2.27}$$

where y_t is the quarterly change in output price, x_2 is labor costs, x_3 is materials input prices, and x_4 is a proxy variable for demand, constructed from the ratio of finished inventory to sales. The standard errors of the estimates are in parentheses.

The findings of (6.2.27) are somewhat unsettling. The contribution of materials input costs is extremely small, less than 1 percent. Furthermore, the proxy variable has the wrong sign. As the inventory-to-sales ratio increases, one would expect the resulting inventory buildup to exert a downward pressure on prices.

There are many reasons that (6.2.27) can go wrong. For instance, pricing behavior across industries is likely to vary, because input combinations are different, labor markets are not homogeneous, and demand may be more elastic or inelastic in one industry than another. In fact, a modified one-way ANCOVA test for the common intercept and slope coefficients,

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_N, \quad N = 17,$$

using the statistic (6.2.21), has a value of 449.28. That is well above the χ^2 critical value of 92.841 for the 1 percent significance level with 64 ($(N-1)K$) degrees of freedom.

The rejection of the hypothesis of homogeneous price behavior across industries suggests a need to modify the model to allow for heterogeneous behavior across industries. However, previous studies have found that output prices are affected mainly by unit labor and materials input costs, and secondly, if at all,

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

by demand factors. Thus, to account for heterogeneous behavior, one can assume that the relationships among variables are proper, but the coefficients are different across industries. But if these coefficients are treated as fixed and different, this will imply a complicated aggregation problem for the price behavior of the U.S. manufacturing sector (e.g., Theil 1954). On the other hand, if the coefficients are treated as random, with common means, there is no aggregation bias (Zellner 1966). The random-coefficient formulation will provide a microeconomic foundation to aggregation, as well as permit the aggregate-price equation to capture more fully the disaggregated industry behavior. Therefore, Barth et al. (1979) used the Swamy random-coefficient formulation, (6.2.6) and (6.2.7), to reestimate the price equation. Their new estimates, with standard errors in parentheses, are

$$\begin{underset{}{\hat{y} =} \quad \underset{(0.0005)}{-0.0006} \quad + \quad \underset{(0.0432)}{0.3093x_2} \quad + \quad \underset{(0.0577)}{0.2687x_3} \quad - \quad \underset{(0.0101)}{0.0082x_4}.} \tag{6.2.28}$$

The estimated dispersion of these coefficients is

$$\hat{\Delta} = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 \\ 0.0000 & -0.0002 & 0.0000 & -0.0001 \\ & 0.0020 & 0.0003 & 0.0081 \\ & & 0.0320 & 0.0030 \\ & & & 0.0014 \end{bmatrix}. \tag{6.2.29}$$

The results of the Swamy random-coefficient formulation appear more plausible than the previous aggregate price specification [equation (6.2.27), which ignores variation across industries] from several points of view: (1) both labor costs and materials costs are now dominant in determining output prices; (2) the proxy variable for demand has the correct sign, although it plays only a small and insignificant role in the determination of manufacturing prices; and (3) productivity, as captured in the intercept term, appears to be increasing.

This example suggests that one must be careful about drawing conclusions on the basis of aggregate data or pooled estimates that do not allow for individual heterogeneity. Such estimates can be misleading in terms of both the size of coefficients and the significance of variables.

6.3 COEFFICIENTS THAT VARY OVER TIME AND CROSS-SECTIONAL UNITS

6.3.1 The Model

Just as in the variable-intercept models, it is possible to assume that the coefficient of the explanatory variable has a component specific to an individual unit

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

and a component specific to a given time period such that

$$y_{it} = \sum_{k=1}^{K} (\bar{\beta}_k + \alpha_{ki} + \lambda_{kt}) x_{kit} + u_{it}, \qquad \begin{array}{l} i = 1, \ldots, N, \\ t = 1, \ldots, T. \end{array} \tag{6.3.1}$$

Stacking all NT observations, we can rewrite (6.3.1) as

$$\mathbf{y} = X\bar{\boldsymbol{\beta}} + \tilde{X}\boldsymbol{\alpha} + \underset{\sim}{\mathbf{X}}\boldsymbol{\lambda} + \mathbf{u}, \tag{6.3.2}$$

where $\mathbf{y}$, X, $\tilde{X}$, $\mathbf{u}$, and $\boldsymbol{\alpha}$ are defined in Section 6.2,

$$\underset{NT \times TK}{\underset{\sim}{\mathbf{X}}} = \begin{bmatrix} \underset{\sim}{\mathbf{X}}_1 \\ \underset{\sim}{\mathbf{X}}_2 \\ \vdots \\ \underset{\sim}{\mathbf{X}}_N \end{bmatrix}, \quad \underset{T \times TK}{\underset{\sim}{\mathbf{X}}_i} = \begin{bmatrix} \mathbf{x}_{i1}' & & & \mathbf{0}' \\ & \mathbf{x}_{i2}' & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{x}_{iT}' \end{bmatrix},$$

and

$$\underset{KT \times 1}{\boldsymbol{\lambda}} = (\boldsymbol{\lambda}_1', \ldots, \boldsymbol{\lambda}_T')', \; \underset{K \times 1}{\boldsymbol{\lambda}_t} = (\lambda_{1t}, \ldots, \lambda_{Kt})',$$

We can also rewrite (6.3.2) as

$$\begin{aligned} \mathbf{y} = {} & X\bar{\boldsymbol{\beta}} + U_1\boldsymbol{\alpha}_1 + U_2\boldsymbol{\alpha}_2 + \cdots + U_K\boldsymbol{\alpha}_K \\ & + U_{K+1}\boldsymbol{\lambda}_1 + \cdots + U_{2K}\boldsymbol{\lambda}_K + U_{2K+1}\mathbf{u}, \end{aligned} \tag{6.3.3}$$

where

$$\underset{NT \times N}{U_k} = \begin{bmatrix} x_{k11} & & & & \\ \vdots & & & & \mathbf{0} \\ x_{k1T} & & & & \\ & x_{k21} & & & \\ & \vdots & & & \\ & x_{k2T} & & & \\ & & \ddots & & \\ & & & x_{kN1} & \\ \mathbf{0} & & & \vdots & \\ & & & x_{kNT} & \end{bmatrix}, \quad k = 1 \ldots, K,$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

$$
\underset{NT\times T}{U_{K+k}} = \begin{bmatrix}
x_{k11} & & & \mathbf{0} \\
& x_{k12} & & \\
& & \vdots & \\
\mathbf{0} & & & x_{k1T} \\
x_{k21} & & & \mathbf{0} \\
& x_{k22} & & \\
& & \ddots & \\
\mathbf{0} & & & x_{k2T} \\
x_{kN1} & & & \mathbf{0} \\
& \ddots & & \\
\mathbf{0} & & & x_{kNT}
\end{bmatrix}, \quad k = 1.\ldots, K, \tag{6.3.4}
$$

$$
U_{2K+1} = I_{NT},
$$

$$
\underset{N\times 1}{\boldsymbol{\alpha}_k} = (\alpha_{k1}, \ldots, \alpha_{kN})', \quad \underset{T\times 1}{\boldsymbol{\lambda}_k} = (\lambda_{k1}, \ldots, \lambda_{kT})'.
$$

When $\boldsymbol{\alpha}_k$ and $\boldsymbol{\lambda}_k$ as well as $\bar{\boldsymbol{\beta}}$ are considered fixed, it is a fixed-effects model; when $\boldsymbol{\alpha}_k$ and $\boldsymbol{\lambda}_k$ are considered random, with $\bar{\boldsymbol{\beta}}$ fixed, equation (6.3.3) corresponds to the mixed analysis-of-variance (ANOVA) model (Hartley and Rao 1967). Thus, model (6.3.1) and its special case, model (6.2.1), fall within the general ANOVA framework.

6.3.2 Fixed-Coefficient Model

When $\boldsymbol{\alpha}_k$ and $\boldsymbol{\lambda}_k$ are treated as fixed, as mentioned earlier, (6.3.1) can be viewed as a fixed-effects ANOVA model. However, the matrix of explanatory variables is $NT \times (T + N + 1)K$, but its rank is only $(T + N - 1)K$; so we must impose $2K$ independent linear restrictions on the coefficients $\boldsymbol{\alpha}_k$ and $\boldsymbol{\lambda}_k$ for estimation of $\bar{\boldsymbol{\beta}}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\lambda}$. A natural way of imposing the constraints in this case is to let[9]

$$
\sum_{i=1}^{N} \alpha_{ik} = 0, \tag{6.3.5}
$$

and

$$
\sum_{t=1}^{T} \lambda_{kt} = 0, \quad k = 1.\ldots, K. \tag{6.3.6}
$$

[9] We did not impose similar restrictions in Chapter 6, Section 6.2.1 because we did not separate $\boldsymbol{\beta}$ from $\boldsymbol{\alpha}_i$.

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Then the best linear unbiased estimators (BLUEs) of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\lambda}$ are the solutions of

$$\min\ (\mathbf{y} - X\bar{\boldsymbol{\beta}} - \tilde{X}\boldsymbol{\alpha} - \underline{\mathbf{X}}\boldsymbol{\lambda})'(\mathbf{y} - X\bar{\boldsymbol{\beta}} - \tilde{X}\boldsymbol{\alpha} - \underline{\mathbf{X}}\boldsymbol{\lambda}) \tag{6.3.7}$$

subject to (6.3.5) and (6.3.6).

6.3.3 Random-Coefficient Model

When $\boldsymbol{\alpha}_i$ and $\boldsymbol{\lambda}_t$ are treated as random, Hsiao (1974a, 1975) assumes that

$$\begin{aligned} \underset{K\times K}{E\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j'} &= \begin{cases} \Delta & \text{if } i = j, \\ \mathbf{0} & \text{if } i \neq j, \end{cases} \\ \underset{K\times K}{E\boldsymbol{\lambda}_t\boldsymbol{\lambda}_s'} &= \begin{cases} \Lambda & \text{if } t = s, \\ \mathbf{0} & \text{if } t \neq s, \end{cases} \end{aligned} \tag{6.3.8}$$

$$E\boldsymbol{\alpha}_i\boldsymbol{\lambda}_t' = \mathbf{0}, \quad E\boldsymbol{\alpha}_i\mathbf{x}_{it}' = \mathbf{0}, \quad E\boldsymbol{\lambda}_t\mathbf{x}_{it}' = \mathbf{0},$$

and

$$E\mathbf{u}_i\mathbf{u}_j' = \begin{cases} \sigma_u^2 I_T & \text{if } i = j, . \\ \mathbf{0} & \text{if } i \neq j. \end{cases}$$

Then the composite error term,

$$\mathbf{v} = \tilde{X}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\lambda} + \mathbf{u}, \tag{6.3.9}$$

has a variance–covariance matrix

$$\begin{aligned} \Omega = E\mathbf{v}\mathbf{v}' = &\begin{bmatrix} X_1\Delta X_1' & & & \mathbf{0} \\ & X_2\Delta X_2' & & \\ & & \ddots & \\ \mathbf{0} & & & X_N\Delta X_N' \end{bmatrix} \\ &+ \begin{bmatrix} D(X_1\Lambda X_1') & D(X_1\Lambda X_2') & \dots & D(X_1\Lambda X_N') \\ D(X_2\Lambda X_1') & D(X_2\Lambda X_2') & & \\ & & & \ddots \\ D(X_N\Lambda X_1') & & & D(X_N\Lambda X_N') \end{bmatrix} + \sigma_u^2 I_{NT}, \end{aligned} \tag{6.3.10}$$

where

$$\underset{T\times T}{D(X_i\Lambda X_j')} = \begin{bmatrix} \mathbf{x}_{i1}'\Lambda\mathbf{x}_{j1} & & & \mathbf{0} \\ & \mathbf{x}_{i2}'\Lambda\mathbf{x}_{j2} & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{x}_{iT}'\Lambda\mathbf{x}_{jT} \end{bmatrix}.$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

We can estimate $\bar{\boldsymbol{\beta}}$ by the least-squares method, but as discussed in Section 6.2.2.2, it is not efficient. Moreover, the conventional formula for the covariance matrix of the least-squares estimator is misleading. If Ω is known, the BLUE of $\bar{\boldsymbol{\beta}}$ is the GLS estimator,

$$\hat{\bar{\boldsymbol{\beta}}}_{\text{GLS}} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\mathbf{y}). \tag{6.3.11}$$

The variance–covariance matrix of the GLS estimator is

$$\text{Var}\,(\hat{\bar{\boldsymbol{\beta}}}_{\text{GLS}}) = (X'\Omega^{-1}X)^{-1}. \tag{6.3.12}$$

Without knowledge of Ω, we can estimate $\bar{\boldsymbol{\beta}}$ and Ω simultaneously by the maximum-likelihood method. However, because of the computational difficulty, a natural alternative is to first estimate Ω, and then substitute the estimated Ω in (6.3.11).

When Δ and Λ are diagonal, it is easy to see from (6.3.3) that Ω is a linear combination of known matrices with unknown weights. So the problem of estimating the unknown covariance matrix is actually the problem of estimating the variance components. Statistical methods developed for estimating the variance (and covariance) components can be applied here (e.g., Anderson 1969, 1970; Rao 1970, 1972). In this section we shall describe only a method due to Hildreth and Houck (1968).[10]

Consider the time-series equation for the ith individual,

$$\mathbf{y}_i = X_i(\bar{\boldsymbol{\beta}} + \boldsymbol{\alpha}_i) + \underline{\mathbf{X}}_i\boldsymbol{\lambda} + \mathbf{u}_i. \tag{6.3.13}$$

We can treat $\boldsymbol{\alpha}_i$ as if it is a vector of constants. Then (6.3.13) is a linear model with heteroscedastic variance. The variance of the error term $r_{it} = \sum_{k=1}^{K}\lambda_{kt}x_{kit} + u_{it}$ is

$$\theta_{it} = E[r_{it}^2] = \sum_{k=1}^{K}\sigma_{\lambda k}^2 x_{kit}^2 + \sigma_u^2. \tag{6.3.14}$$

Let $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{iT})'$; then

$$\boldsymbol{\theta}_i = \dot{X}_i\boldsymbol{\sigma}_\lambda^2, \tag{6.3.15}$$

where the first element of $\mathbf{x}_{it} = 1$, $\dot{X}_i$ is X_i with each of its elements squared, and $\boldsymbol{\sigma}_\lambda^2 = (\sigma_{\lambda 1}^2 + \sigma_u^2, \sigma_{\lambda 2}^2, \ldots, \sigma_{\lambda K}^2)'$.

An estimate of $\mathbf{r}_i$ can be obtained as the least-squares residual, $\hat{\mathbf{r}}_i = \mathbf{y}_i - X_i\hat{\boldsymbol{\beta}}_i = M_i\mathbf{y}_i$, where $\hat{\boldsymbol{\beta}}_i = (X_i'X_i)^{-1}X_i'\mathbf{y}_i$ and $M_i = I_T - X_i(X_i'X_i)^{-1}X_i'$. Squaring each element of $\hat{\mathbf{r}}_i$ and denoting it by $\dot{\hat{\mathbf{r}}}_i$, we have

$$E(\dot{\hat{\mathbf{r}}}_i) = \dot{M}_i\theta_i = F_i\sigma_\lambda^2, \tag{6.3.16}$$

where $\dot{M}_i$ is M_i with each of its elements squared, and $F_i = \dot{M}_i\dot{X}_i$.

[10] It has been shown (Hsiao 1975) that the Hildreth–Houck estimator is the minimum-norm quadratic unbiased estimator of Rao (1970).

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Repeating the foregoing process for all i gives

$$E(\dot{\hat{\mathbf{r}}}) = F\boldsymbol{\sigma}_\lambda^2, \tag{6.3.17}$$

where $\dot{\hat{\mathbf{r}}} = (\dot{\hat{\mathbf{r}}}_1, \ldots, \dot{\hat{\mathbf{r}}}_N)'$, and $F = (F_1', \ldots, F_N')'$. Application of least-squares to (6.3.17) yields a consistent estimator of σ_λ^2,

$$\hat{\boldsymbol{\sigma}}_\lambda^2 = (F'F)^{-1}F'\dot{\hat{\mathbf{r}}}. \tag{6.3.18}$$

Similarly, we can apply the same procedure with respect to each time period to yield a consistent estimator of $\boldsymbol{\sigma}_\alpha^2 = (\sigma_{\alpha_1}^2 + \sigma_u^2, \sigma_{\alpha_2}^2, \ldots, \sigma_{\alpha_K}^2)'$. To obtain separate estimates of σ_u^2, $\sigma_{\alpha_1}^2$, and $\sigma_{\lambda 1}^2$, we note that $E(\mathbf{x}_{it}'\boldsymbol{\alpha}_i + u_{it})(\mathbf{x}_{it}'\boldsymbol{\lambda}_t + u_{it}) = \sigma_u^2$. So, letting $\hat{s}_{it}$ denote the residual obtained by applying least-squares separately to each time period, we can consistently estimate σ_u^2 by

$$\hat{\sigma}_u^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{r}_{it}\hat{s}_{it}. \tag{6.3.19}$$

Subtracting (6.3.19) from an estimated $\sigma_{\alpha_1}^2 + \sigma_u^2$ and $\sigma_{\lambda_1}^2 + \sigma_u^2$, we obtain consistent estimates of $\sigma_{\alpha_1}^2$ and $\sigma_{\lambda_1}^2$, respectively.

Substituting consistently estimated values of $\boldsymbol{\sigma}_\alpha^2$, $\boldsymbol{\sigma}_\lambda^2$, and σ_u^2 into (6.3.11), one can show that when N and T both tend to infinity and N/T tends to a nonzero constant, the two-stage Aitken estimator is asymptotically as efficient as if one knew the true Ω. Also, Kelejian and Stephan (1983) have pointed out that contrary to the conventional regression model, the speed of convergence of $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ here is not $(NT)^{1/2}$, but max $(N^{1/2}, T^{1/2})$.

If one is interested in predicting the random components associated with an individual, Lee and Griffiths (1979) have shown that the predictor

$$\hat{\boldsymbol{\alpha}} = (I_N \otimes \Delta)X'\Omega^{-1}(\mathbf{y} - X\hat{\boldsymbol{\beta}}_{\text{GLS}}) \tag{6.3.20}$$

is the BLUE.

To test for the random variation of the coefficients, we can again apply the Breusch and Pagan (1979) Lagrangian-multiplier test for heteroscedasticity. Because for given i, $\boldsymbol{\alpha}_i$ is fixed, the error term $\mathbf{x}_{it}'\boldsymbol{\lambda}_t + u_{it}$ will be homoscedastic if the coefficients are not varying over time. Therefore, under the null, one-half the explained sum of squares in a regression[11]

$$\begin{aligned}\frac{\hat{u}_{it}^2}{\hat{\sigma}_u^2} &= \dot{\mathbf{x}}_{it}'\boldsymbol{\sigma}_\lambda^2 + \epsilon_{it}, \qquad i = 1, \ldots, N,\\ &\qquad t = 1, \ldots, T,\end{aligned} \tag{6.3.21}$$

[11] Let $(y_{it} - y)$ be the deviation of the sample mean, and let $(\widehat{y_{it} - y})$ be its least-squares prediction. Then the explained sum of squares is $\Sigma(\widehat{y_{it} - y})^2$.

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

is distributed asymptotically as χ^2 with $K-1$ degrees of freedom, where $\hat{u}_{it} = y_{it} - \hat{\boldsymbol{\beta}}_i' \mathbf{x}_{it}$, $\hat{\sigma}_u^2 = \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \hat{\boldsymbol{\beta}}_i' \mathbf{x}_{it})^2 / NT$, and $\dot{\mathbf{x}}_{it}$ is $\mathbf{x}_{it}$ with each element squared.[12]

Similarly, we can test for random variation across cross-sectional units by regressing

$$\frac{\hat{u}_{it}^{*2}}{\hat{\sigma}_u^{*2}} = \dot{\mathbf{x}}_{it}' \boldsymbol{\sigma}_\alpha^2 + \epsilon_{it}^*, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{6.3.22}$$

where $\hat{u}_{it}^* = y_{it} - \hat{\boldsymbol{\beta}}_t' \mathbf{x}_{it}$, $\hat{\sigma}_u^{*2} = \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^{*2} / NT$, and $\hat{\boldsymbol{\beta}}_t$ is the least-squares estimate of $\boldsymbol{\beta}_t = \boldsymbol{\beta} + \boldsymbol{\lambda}_t$ across cross-sectional units for a given t. Under the null hypothesis of no random variation across cross-sectional units, one-half of the explained sum of squares of (6.3.22) is asymptotically χ^2 distributed with $K-1$ degrees of freedom.

We can also test the random variation indirectly by applying the classic ANCOVA test. For details, see Hsiao (1974a).

Swamy and Mehta (1977) have proposed a more general type of time-varying-component model by allowing $E\boldsymbol{\lambda}_t \boldsymbol{\lambda}_t' = \Lambda_t$ to vary over t. However, models with the coefficients varying randomly across cross-sectional units and over time have not gained much acceptance in empirical investigations. Part of the reason is because the inversion of Ω is at least of order max (NK, TK) (Hsiao 1974a). For any panel data of reasonable size, this would be a computationally demanding problem.

6.4 COEFFICIENTS THAT EVOLVE OVER TIME

6.4.1 The Model

There is a large amount of empirical evidence that parameters of a model change over time. For instance, financial liberalization or changes in monetary policy can cause the relationships between economic variables to alter. If a constant-parameter model is used, misspecification may occur. On the other hand, if a model is too flexible in its treatment of parameter change, over-fitting or imprecise inferences can occur. In this section, we discuss some commonly used time-varying-parameter models that entail a smooth evolution.[13]

In most models with coefficients evolving over time it is assumed that there is no individual heterogeneity (e.g., Zellner, Hong, and Min 1991). At a given t, the coefficient vectors $\boldsymbol{\beta}_t$ are identical for all cross-sectional units. For this reason we shall discuss the main issues of time-varying-parameter models assuming that $N = 1$, and then indicate how this analysis can be modified when $N > 1$.

[12] Note here that the first term $\dot{x}_{1it} = 1$. So the null hypothesis is $(\sigma_{\lambda 2}^2, \ldots, \sigma_{\lambda K}^2) = (0, \ldots, 0)$.

[13] This section is largely drawn from the work of Chow (1983, their Chapter 10).

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

As shown by Chow (1983, Chapter 10), a wide variety of time-varying-parameter models can be put in the general form

$$y_t = \boldsymbol{\beta}_t' \mathbf{x}_t + u_t, \tag{6.4.1}$$

and

$$\boldsymbol{\beta}_t = H\boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \quad t = 1\ldots, T, \tag{6.4.2}$$

where $\mathbf{x}_t$ is a $K \times 1$ vector of exogenous variables; u_t is independent normal, with mean 0 and variance σ_u^2; $\boldsymbol{\eta}_t$ is a K-variant independent normal random variable, with mean 0 and covariance matrix Ψ; and $\boldsymbol{\eta}$ and u are independent. For instance, when $H = I_K$, it is the random-walk model of Cooley and Prescott (1976). When $H = I_K$ and $\Psi = \mathbf{0}$, this model is reduced to the standard regression model.

The Rosenberg (1972, 1973) return-to-normality model can also be put into this form. The model corresponds to replacing $\boldsymbol{\beta}_t$ and $\boldsymbol{\beta}_{t-1}$ in (6.4.2) by $(\boldsymbol{\beta}_t - \boldsymbol{\beta})$ and $(\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta})$ and restricting the absolute value of the characteristic roots of H to <1. Although this somewhat changes the formulation, if we define $\boldsymbol{\beta}_t^* = \boldsymbol{\beta}_t - \boldsymbol{\beta}$ and $\bar{\boldsymbol{\beta}}_t = \boldsymbol{\beta}$, the return-to-normality model can be rewritten as

$$\begin{aligned} y_t &= (\mathbf{x}_t', \mathbf{x}_t') \begin{bmatrix} \bar{\boldsymbol{\beta}}_t \\ \boldsymbol{\beta}_t^* \end{bmatrix} + u_t \\ \begin{bmatrix} \bar{\boldsymbol{\beta}}_t \\ \boldsymbol{\beta}_t^* \end{bmatrix} &= \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & H \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{\beta}}_{t-1} \\ \boldsymbol{\beta}_{t-1}^* \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\eta}_t \end{bmatrix}, \end{aligned} \tag{6.4.3}$$

which is a special case of (6.4.1) and (6.4.2).

Similarly, we can allow $\boldsymbol{\beta}_t$ to be stationary, with constant mean $\boldsymbol{\beta}$ (Pagan 1980). Suppose

$$\begin{aligned} y_t &= \mathbf{x}_t'\boldsymbol{\beta} + \mathbf{x}_t'\boldsymbol{\beta}_t^* + u_t, \\ \boldsymbol{\beta}_t^* &= \boldsymbol{\beta}_t - \boldsymbol{\beta} = A^{-1}(\mathcal{L})\boldsymbol{\epsilon}_t, \end{aligned} \tag{6.4.4}$$

where $A(\mathcal{L})$ is a ratio of polynomials of orders p and q in the lag operator $\mathcal{L}(\mathcal{L}\boldsymbol{\epsilon}_t = \boldsymbol{\epsilon}_{t-1})$, and $\boldsymbol{\epsilon}$ is independent normal, so that $\boldsymbol{\beta}_t^*$ follows an autoregressive moving-average (ARMA) (p, q) process. Because an ARMA of order p and q can be written as a first-order autoregressive process, this model can again be put in the form of (6.4.1) and (6.4.2). For example,

$$\boldsymbol{\beta}_t^* = B_1\boldsymbol{\beta}_{t-1}^* + B_2\boldsymbol{\beta}_{t-2}^* + \boldsymbol{\epsilon}_t + B_3\boldsymbol{\epsilon}_{t-1} \tag{6.4.5}$$

can be written as

$$\tilde{\boldsymbol{\beta}}_t^* = \begin{bmatrix} \boldsymbol{\beta}_t^* \\ \boldsymbol{\beta}_{t-1}^* \\ \boldsymbol{\epsilon}_t \end{bmatrix} = \begin{bmatrix} B_1 & B_2 & B_3 \\ I & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{t-1}^* \\ \boldsymbol{\beta}_{t-2}^* \\ \boldsymbol{\epsilon}_{t-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_t \\ \mathbf{0} \\ \boldsymbol{\epsilon}_t \end{bmatrix} = H\tilde{\boldsymbol{\beta}}_{t-1}^* + \boldsymbol{\eta}_t. \tag{6.4.6}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Thus, we can write Pagan's model in the form

$$y_t = (\mathbf{x}_t', \tilde{\mathbf{x}}_t') \begin{bmatrix} \bar{\boldsymbol{\beta}}_t \\ \tilde{\boldsymbol{\beta}}_t^* \end{bmatrix} + u_t, \tag{6.4.4a}$$

where $\tilde{\mathbf{x}}_t' = (\mathbf{x}_t', \mathbf{0}', \mathbf{0}')$. Equation (6.4.4a) is then formally equivalent to (6.4.3).

The Kalman filter (Kalman 1960) provides a basis for computing the maximum-likelihood estimators and predicting the evolution of the time path of $\boldsymbol{\beta}_t$ for this type of the model. In this section we first consider the problem of estimating $\boldsymbol{\beta}_t$ using information $\mathcal{I}_s$, up to the time s, assuming that σ_u^2, Ψ, and H are known. We denote the conditional expectation of $\boldsymbol{\beta}_t$, given $\mathcal{I}_s$, as $E(\boldsymbol{\beta}_t \mid \mathcal{I}_s) = \boldsymbol{\beta}_{t|s}$. The evaluation of $\boldsymbol{\beta}_{t|s}$ is called filtering when $t = s$; it is called smoothing when $s > t$; it is called prediction when $s < t$. We then study the problem of estimating σ_u^2, Ψ, and H by the method of maximum likelihood. Finally, we consider the problem of testing for constancy of the parameters.

6.4.2 Predicting $\boldsymbol{\beta}_t$ by the Kalman Filter

Denote $(y_1, \ldots, y_t)$ by Y_t. By definition, the conditional mean of $\boldsymbol{\beta}_t$, given Y_t, is

$$\begin{aligned} \boldsymbol{\beta}_{t|t} &= E(\boldsymbol{\beta}_t \mid y_t, Y_{t-1}) \\ &= E(\boldsymbol{\beta}_t \mid Y_{t-1}) + L_t[y_t - E(y_t \mid Y_{t-1})] \\ &= \boldsymbol{\beta}_{t|t-1} + L_t[y_t - \mathbf{x}_t'\boldsymbol{\beta}_{t|t-1}]. \end{aligned} \tag{6.4.7}$$

where $y_t - E(y_t \mid Y_{t-1})$ denotes the additional information of y_t not contained in Y_{t-1} and L_t denotes the adjustment factor of $\boldsymbol{\beta}_{t|t-1}$ because of this additional information. If L_t is known, (6.4.7) can be used to update our estimate $\boldsymbol{\beta}_{t|t-1}$ to form $\boldsymbol{\beta}_{t|t}$.

To derive L_t, we know from our assumption on $\boldsymbol{\eta}_t$ and u_t that, conditional on $\mathbf{x}_t$, y_t and $\boldsymbol{\beta}_t$ are jointly normally distributed. The normal-distribution theory (Anderson 1985, Chapter 2) states that, conditional on Y_{t-1} (and X_t), the mean of $\boldsymbol{\beta}_t$, given y_t is $E(\boldsymbol{\beta}_t \mid Y_{t-1}) + \text{Cov}(\boldsymbol{\beta}_t, y_t \mid Y_{t-1})\text{Var}(y_t \mid Y_{t-1})^{-1}[y_t - E(y_t \mid Y_{t-1})]$. Therefore,

$$L_t = [E(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t-1})(y_t - y_{t|t-1})]\text{Var}(y_t \mid Y_{t-1})^{-1}, \tag{6.4.8}$$

where $y_{t|t-1} = E(y_t \mid Y_{t-1}) = \mathbf{x}_t'\boldsymbol{\beta}_{t|t-1}$. Denoting the covariance matrix $\text{Cov}(\boldsymbol{\beta}_t \mid Y_{t-1}) = E(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t-1})(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t-1})'$ by $\Sigma_{t|t-1}$, we have

$$\begin{aligned} &E(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t-1})(y_t - y_{t|t-1}) \\ &\quad = E\{(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t-1})[(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t-1})'\mathbf{x}_t + u_t]\} = \Sigma_{t|t-1}\mathbf{x}_t, \end{aligned} \tag{6.4.9}$$

and

$$\begin{aligned} \text{Var}(y_t \mid Y_{t-1}) &= E[\mathbf{x}_t'(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t-1}) + u_t][(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t-1})'\mathbf{x}_t + u_t] \\ &= \mathbf{x}_t'\Sigma_{t|t-1}\mathbf{x}_t + \sigma_u^2. \end{aligned} \tag{6.4.10}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Hence, (6.4.8) becomes

$$L_t = \Sigma_{t|t-1}\mathbf{x}_t(\mathbf{x}_t'\Sigma_{t|t-1}\mathbf{x}_t + \sigma_u^2)^{-1}. \tag{6.4.11}$$

From (6.4.2) we have

$$\boldsymbol{\beta}_{t|t-1} = H\boldsymbol{\beta}_{t-1|t-1}. \tag{6.4.12}$$

Thus, we can compute $\Sigma_{t|t-1}$ recursively by

$$\begin{aligned}\Sigma_{t|t-1} &= E(\boldsymbol{\beta}_t - H\boldsymbol{\beta}_{t-1|t-1})(\boldsymbol{\beta}_t - H\boldsymbol{\beta}_{t-1|t-1})' \\ &= E[H(\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}_{t-1|t-1}) + \boldsymbol{\eta}_t] \\ &\quad \cdot [H(\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}_{t-1|t-1}) + \boldsymbol{\eta}_t]' \\ &= H\Sigma_{t-1|t-1}H' + \Psi.\end{aligned} \tag{6.4.13}$$

Next, from (6.4.1) and (6.4.7) we can write

$$\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t} = \boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t-1} - L_t[\mathbf{x}_t'(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t-1}) + u_t]. \tag{6.4.14}$$

Taking the expectation of the product of (6.4.14) and itstranspose, and using (6.4.11), we obtain

$$\begin{aligned}\Sigma_{t|t} &= \Sigma_{t|t-1} - L_t(\mathbf{x}_t'\Sigma_{t|t-1}\mathbf{x}_t + \sigma_u^2)L_t' \\ &= \Sigma_{t|t-1} - \Sigma_{t|t-1}\mathbf{x}_t(\mathbf{x}_t'\Sigma_{t|t-1}\mathbf{x}_t + \sigma_u^2)^{-1}\mathbf{x}_t'\Sigma_{t|t-1}.\end{aligned} \tag{6.4.15}$$

Equations (6.4.13) and (6.4.15) can be used to compute $\Sigma_{t|t}(t = 1, 2, \ldots)$ successively, given $\Sigma_{0|0}$. Having computed $\Sigma_{t|t-1}$, we can use (6.4.11) to compute L_t. Given L_t, (6.4.7) and (6.4.12) can be used to compute $\boldsymbol{\beta}_{t|t}$ from $\boldsymbol{\beta}_{t-1|t-1}$ if $\boldsymbol{\beta}_{0|0}$ is known.

Similarly, we can predict $\boldsymbol{\beta}_t$ using future observations $y_{t+1}, y_{t+2}, \ldots, y_{t+n}$. We first consider the regression of $\boldsymbol{\beta}_t$ on y_{t+1}, conditional on Y_t. Analogous to (6.4.7) and (6.4.11) are

$$\boldsymbol{\beta}_{t|t+1} = \boldsymbol{\beta}_{t|t} + F_{t|t+1}(y_{t+1} - y_{t+1|t}) \tag{6.4.16}$$

and

$$F_{t|t+1} = [E(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t})(y_{t+1} - y_{t+1|t})'][\mathrm{Cov}(y_{t+1} \mid Y_t)]^{-1}. \tag{6.4.17}$$

To derive the matrix $F_{t|t+1}$ of regression coefficients, we use (6.4.1) and (6.4.2) to write

$$\begin{aligned}y_{t+1} - y_{t+1|t} &= \mathbf{x}_{t+1}'(\boldsymbol{\beta}_{t+1} - \boldsymbol{\beta}_{t+1|t}) + u_{t+1} \\ &= \mathbf{x}_{t+1}'H(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t}) + \mathbf{x}_{t+1}'\boldsymbol{\eta}_{t+1} + u_{t+1}.\end{aligned} \tag{6.4.18}$$

Combining (6.4.17), (6.4.18), (6.4.10), and (6.4.11), we have

$$\begin{aligned}F_{t|t+1} &= \Sigma_{t|t}H'\mathbf{x}_{t+1}(\mathbf{x}_{t+1}'\Sigma_{t+1|t}\mathbf{x}_{t+1} + \sigma_u^2)^{-1} \\ &= \Sigma_{t|t}H'\Sigma_{t+1|t}^{-1}L_{t+1}.\end{aligned} \tag{6.4.19}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Therefore, from (6.4.19) and (6.4.14), we can rewrite (6.4.16) as

$$\boldsymbol{\beta}_{t|t+1} = \boldsymbol{\beta}_{t|t} + \Sigma_{t|t} H' \Sigma_{t+1|t}^{-1} (\boldsymbol{\beta}_{t+1|t+1} - \boldsymbol{\beta}_{t+1|t}). \tag{6.4.20}$$

Equation (6.4.20) can be generalized to predict $\boldsymbol{\beta}_t$ using future observations $y_{t+1}, \ldots, y_{t+n}$,

$$\boldsymbol{\beta}_{t|t+n} = \boldsymbol{\beta}_{t|t+n-1} + F_t^* (\boldsymbol{\beta}_{t+1|t+n} - \boldsymbol{\beta}_{t+1|t+n-1}), \tag{6.4.21}$$

where $F_t^* = \Sigma_{t|t} H' \Sigma_{t+1|t}^{-1}$. The proof of this is given by Chow (1983, Chapter 10).

When H, Ψ, and σ_u^2 are known, (6.4.7) and (6.4.21) trace out the time path of $\boldsymbol{\beta}_t$ and provide the minimum-mean-square-error forecast of the future values of the dependent variable, given the initial values $\boldsymbol{\beta}_{0|0}$ and $\Sigma_{0|0}$. To obtain the initial values of $\boldsymbol{\beta}_{0|0}$ and $\Sigma_{0|0}$, Sant (1977) suggested using the GLS method on the first K observations of y_t and $\mathbf{x}_t$. Noting that

$$\begin{aligned} \boldsymbol{\beta}_t &= H\boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t \\ &= H^2 \boldsymbol{\beta}_{t-2} + \boldsymbol{\eta}_t + H\boldsymbol{\eta}_{t-1} \\ &= H^{t-j} \boldsymbol{\beta}_j + \boldsymbol{\eta}_t + H\boldsymbol{\eta}_{t-1} + \cdots + H^{t-j-1} \boldsymbol{\eta}_j, \end{aligned} \tag{6.4.22}$$

and assuming that H^{-1} exists, we can also write y_k in the form

$$\begin{aligned} y_k &= \mathbf{x}_k' \boldsymbol{\beta}_k + u_k \\ &= \mathbf{x}_k' [H^{-K+k} \boldsymbol{\beta}_K - H^{-K+k} \boldsymbol{\eta}_K - \cdots - H^{-1} \boldsymbol{\eta}_{k+1}] + u_k. \end{aligned}$$

Thus, $(y_1, \ldots, y_K)$ can be written as

$$\begin{aligned} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_K \end{bmatrix} &= \begin{bmatrix} \mathbf{x}_1' H^{-K+1} \\ \mathbf{x}_2' H^{-K+2} \\ \vdots \\ \mathbf{x}_K' \end{bmatrix} \boldsymbol{\beta}_K + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_K \end{bmatrix} \\ &\quad - \begin{bmatrix} \mathbf{x}_1' H^{-1} & \mathbf{x}_1' H^{-2} & \cdots & \mathbf{x}_1' H^{-K+1} \\ \mathbf{0}' & \mathbf{x}_2' H^{-1} & \cdots & \mathbf{x}_2' H^{-K+2} \\ & & \cdots & \cdot \\ & & & \mathbf{x}_{K-1}' H^{-1} \\ & & & \mathbf{0}' \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_2 \\ \boldsymbol{\eta}_3 \\ \vdots \\ \boldsymbol{\eta}_K \end{bmatrix}. \end{aligned} \tag{6.4.23}$$

Applying GLS to (6.4.23) gives

$$\begin{aligned} \Sigma_{K|K} = \sigma_u^2 \{ &[H'^{-K+1} \mathbf{x}_1, H'^{-K+2} \mathbf{x}_2, \ldots, \mathbf{x}_K] \\ &\cdot [I_K + A_K (I_{K-1} \otimes P) A_K']^{-1} [H^{-K+1} \mathbf{x}_1, \ldots, \mathbf{x}_K]' \}^{-1} \end{aligned} \tag{6.4.24}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

and

$$\boldsymbol{\beta}_{K|K} = \frac{1}{\sigma_u^2}\Sigma_{K|K}[H'^{-K+1}\mathbf{x}_1, H'^{-K+2}\mathbf{x}_2, \ldots, \mathbf{x}_K]$$
$$[I_K + A_K(I_{K-1} \otimes P)A_K']^{-1}\begin{bmatrix} y_1 \\ \vdots \\ y_K \end{bmatrix}, \tag{6.4.25}$$

where $P = \sigma_u^{-2}\Psi$, and A_K is the coefficient matrix of $(\boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_K)'$ in (6.4.23). The initial estimators, $\boldsymbol{\beta}_{K|K}$ and $\Sigma_{K|K}$, are functions of σ_u^2, Ψ, and H.

6.4.3 Maximum-Likelihood Estimation

When H, Ψ, and σ_u^2 are unknown, (6.4.7) opens the way for maximum-likelihood estimation without the need for repeated inversions of covariance matrices of large dimensions. To form the likelihood function, we note that

$$y_t - y_{t|t-1} = \mathbf{x}_t'(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t-1}) + u_t = y_t - x_t'\boldsymbol{\beta}_{t|t-1} \tag{6.4.26}$$

is normal and serially uncorrelated. Hence, the joint density of $(y_1, \ldots, y_T)$ can be written as the product of the conditional density of $(y_{K+1}, \ldots, y_T \mid y_1, \ldots, y_K)$ and the marginal density of $(y_1, \ldots, y_K)$. The log-likelihood function of $(y_{K+1}, \ldots, y_T)$, given $(y_1, \ldots, y_K)$, is

$$\log L = -\frac{T-K}{2}\log 2\pi - \frac{1}{2}\sum_{t=K+1}^{T}\log(\mathbf{x}_t'\Sigma_{t|t-1}\mathbf{x}_t + \sigma_u^2)$$
$$-\frac{1}{2}\sum_{t=K+1}^{T}\frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta}_{t|t-1})^2}{\mathbf{x}_t'\Sigma_{t|t-1}\mathbf{x}_t + \sigma_u^2}. \tag{6.4.27}$$

The first K observations are used to compute $\Sigma_{K|K}$ and $\boldsymbol{\beta}_{K|K}$ [equations (6.4.24) and (6.4.25)] as functions of σ_u^2, Ψ, and H. Hence, the data $\boldsymbol{\beta}_{t|t-1}$ and $\Sigma_{t|t-1}$ $(t = K+1, \ldots, T)$ required to evaluate $\log L$ are functions of σ_u^2, Ψ, and H, as given by (6.4.13), (6.4.15), (6.4.12), and (6.4.11). To find the maximum of (6.4.27), numerical methods will have to be used.

When we estimate the model (6.4.1) and (6.4.2) using panel data, all the derivations in Section 6.4.2 remain valid if we replace y_t, $\mathbf{x}_t$, $\mathbf{u}_t$, and σ_u^2 by the $N \times 1$ vector $\mathbf{y}_t = (y_{1t}, \ldots, y_{Nt})'$, the $N \times K$ matrix $X_t = (\mathbf{x}_{1t}, \ldots, \mathbf{x}_{Nt})'$, the $N \times 1$ vector $\mathbf{u}_t = (u_{1t}, \ldots, u_{Nt})'$, and $\sigma_u^2 I_N$ in appropriate places. The MLE can be carried out in the same way as outlined in this section, except that

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

the likelihood function (6.4.27) is replaced by

$$\begin{aligned}\text{Log } L = \text{const } &- \frac{1}{2}\sum_t \log \mid X_t'\Sigma_{t|t-1}X_t + \sigma_u^2 I_N \mid \\ &- \frac{1}{2}\sum_t (\mathbf{y}_t - X_t\boldsymbol{\beta}_{t|t-1})' \\ &\cdot (X_t\Sigma_{t|t-1}X_t' + \sigma_u^2 I_N)^{-1}(\mathbf{y}_t - X_t\boldsymbol{\beta}_{t|t-1}).\end{aligned} \tag{6.4.27′}$$

However, we no longer need to use the first K period observations to start the iteration. If $N > K$, we need to use only the first-period cross-sectional data to obtain $\boldsymbol{\beta}_{1|1}$ and $\Sigma_{1|1}$. Additional details with regard to the computation can be found in Harvey (1978) and Harvey and Phillips (1982).

6.4.4 Tests for Parameter Constancy

A simple alternative to the null hypothesis of constancy of regression coefficients over time is

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \tag{6.4.28}$$

where $\boldsymbol{\eta}_t$ is assumed independently normally distributed, with mean 0 and a diagonal covariance matrix Ψ. Regarding $\boldsymbol{\beta}_0$ as fixed, we have

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_0 + \sum_{s=1}^{t} \boldsymbol{\eta}_s. \tag{6.4.29}$$

Thus, the regression model becomes

$$\begin{aligned} y_t = \mathbf{x}_t'\boldsymbol{\beta}_t + u_t &= \mathbf{x}_t'\boldsymbol{\beta}_0 + u_t + \mathbf{x}_t'\left(\sum_{s=1}^{t}\boldsymbol{\eta}_s\right) \\ &= \mathbf{x}_t'\boldsymbol{\beta}_0 + u_t^*, \end{aligned} \tag{6.4.30}$$

where $u_t^* = u_t + \mathbf{x}_t'(\sum_{s=1}^{t}\boldsymbol{\eta}_s)$ has variance

$$Eu_t^{*2} = \sigma_u^2 + t\mathbf{x}_t'\Psi\mathbf{x}_t. \tag{6.4.31}$$

For $\Psi = \text{diag}\{\psi_{kk}\}$, (6.4.31) becomes

$$Eu_t^{*2} = \sigma_u^2 + t\sum_{k=1}^{K} x_{kt}^2\psi_{kk}, \quad t = 1\ldots., T. \tag{6.4.32}$$

The null hypothesis states that $\Psi = \mathbf{0}$. Hence, the Breusch and Pagan (1979) Lagrangian-multiplier test applied here is to regress $\hat{u}_t^2/\hat{\sigma}_u^2$ on $t(1, x_{2t}^2, \ldots, x_{Kt}^2)$, $t = 1, \ldots, T$, where $\hat{u}_t$ is the least-squares residual $\hat{u}_t = y_t - \hat{\boldsymbol{\beta}}'\mathbf{x}_t$, $\hat{\boldsymbol{\beta}} = (\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t')^{-1}(\sum_{t=1}^{T}\mathbf{x}_t y_t)$, and $\hat{\sigma}_u^2 = \sum_{t=1}^{T}\hat{u}_t^2/T$. Under the

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

null hypothesis, one-half the explained sum of squares of this regression is asymptotically χ^2 distributed, with K degrees of freedom.[14]

When panel data are available, it is possible to test for parameter constancy indirectly using the classic ANCOVA test. By the assumption that the parameter vector $\boldsymbol{\beta}_t$, is constant over cross-sectional units in the same period, an indirect test is to postulate the null hypothesis,

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_T = \boldsymbol{\beta}.$$

If the disturbances of the regression model $y_{it} = \boldsymbol{\beta}_t' \mathbf{x}_{it} + u_{it}$ are independently normally distributed over i and t, then the test statistic F_3' from Chapter 2 has an F distribution with $(T-1)K$ and $N(T-K)$ degrees of freedom under the null.

If the null hypothesis is rejected, we can use the information that under mild regularity conditions $\text{plim}_{N\to\infty}\hat{\boldsymbol{\beta}}_t = \boldsymbol{\beta}_t$, $t = 1, \ldots, T$, to investigate the nature of variation in the parameters over time. We can apply the Box–Jenkins (1970) method on $\hat{\boldsymbol{\beta}}_t$ to identify a suitable stochastic process with which to model the parameter variation.

6.5 COEFFICIENTS THAT ARE FUNCTIONS OF OTHER EXOGENOUS VARIABLES

Sometimes, instead of assuming that parameters are random draws from a common distribution, an investigation of possible dependence of $\boldsymbol{\beta}_{it}$ on characteristics of the "individuals" or "time" is of considerable interest (e.g., Amemiya 1978b; Hendricks, Koenker, and Poirier 1979; Singh et al. 1976; Swamy and Tinsley 1977; Wachter 1970). A general formulation of stochastic-parameter models with systematic components can be expressed within the context of the linear model. Suppose that

$$\mathbf{y}_i = X_{i1}\boldsymbol{\beta}_1 + X_{i2}\boldsymbol{\beta}_{2i} + \mathbf{u}_i, \quad i = 1\ldots., N, \tag{6.5.1}$$

and

$$\boldsymbol{\beta}_{2i} = Z_i\boldsymbol{\gamma} + \boldsymbol{\eta}_{2i} \tag{6.5.2}$$

where X_{i1} and X_{i2} denote the $T \times K_1$ and $T \times K_2$ matrices of the time-series observations of the first K_1 and last $K_2 (= K - K_1)$ exogenous variables for the ith individual, $\boldsymbol{\beta}_1$ is a $K_1 \times 1$ vector of fixed constants, $\boldsymbol{\beta}_{2i}$ is a $K_2 \times 1$ vector that varies according to (6.5.2); Z_i and $\boldsymbol{\gamma}$ are a $K_2 \times M$ matrix of known constants and a $M \times 1$ vector of unknown constants, respectively; and $\mathbf{u}_i$ and $\boldsymbol{\eta}_{2i}$ are $T \times 1$ and $K_2 \times 1$ vectors of unobservable random variables that are assumed independent of X_i and Z_i. For example, in Wachter (1970), $\mathbf{y}_i$ is a vector of time-series observations on the logarithm of the relative wage rate in the ith industry. X_{i1} contains the logarithm of such variables as the

[14] Note that under the alternative, u_t^* is serially correlated. Hence, the Breusch and Pagan test may not be powerful against the alternative.

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

relative value-added in the ith industry and the change in the consumer price, X_{i2} consists of a single vector of time series observations on the logarithm of unemployment, and Z_i contains the degree of concentration and the degree of unionization in the ith industry.

For simplicity, we assume that $\mathbf{u}_i$ and $\boldsymbol{\eta}_{2i}$ are uncorrelated with each other and have 0 means. The variance–covariance matrices of $\mathbf{u}_i$ and $\boldsymbol{\eta}_{2i}$ are given by

$$E\mathbf{u}_i\mathbf{u}_j' = \sigma_{ij} I_T \tag{6.5.3}$$

and

$$E\boldsymbol{\eta}_{2i}\boldsymbol{\eta}_{2j}' = \begin{cases} \Lambda & \text{if } i = j, \\ \mathbf{0} & \text{if } i \neq j. \end{cases} \tag{6.5.4}$$

Let $\Sigma = (\sigma_{ij})$. We can write the variance–covariance of $\mathbf{u} = (\mathbf{u}_1', \ldots, \mathbf{u}_N')'$ and $\boldsymbol{\eta}_2 = (\boldsymbol{\eta}_{21}', \ldots, \boldsymbol{\eta}_{2N}')'$ as

$$E\mathbf{u}\mathbf{u}' = \Sigma \otimes I_T \tag{6.5.5}$$

and

$$E\boldsymbol{\eta}_2\boldsymbol{\eta}_2' = \begin{bmatrix} \Lambda & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Lambda \end{bmatrix} = \tilde{\Lambda}. \tag{6.5.6}$$

Combining (6.5.1) and (6.5.2), we have

$$\mathbf{y} = X_1\boldsymbol{\beta}_1 + \mathbf{W}\boldsymbol{\gamma} + \tilde{X}_2\boldsymbol{\eta}_2 + \mathbf{u}, \tag{6.5.7}$$

where

$$\underset{NT\times 1}{\mathbf{y}} = (y_1', \ldots, \mathbf{y}_N')',$$

$$\underset{NT\times K_1}{X_1} = (X_{11}', \ldots, X_{N1}')',$$

$$\underset{NT\times M}{\mathbf{W}} = (Z_1'X_{12}', Z_2'X_{22}', \ldots, Z_N'X_{N2}')',$$

$$\underset{NT\times NK_2}{\tilde{X}_2} = \begin{bmatrix} X_{12} & & & \mathbf{0} \\ & X_{22} & & \\ & & \ddots & \\ \mathbf{0} & & & X_{N2} \end{bmatrix},$$

and

$$\underset{NK_2\times 1}{\boldsymbol{\eta}_2} = (\boldsymbol{\eta}_{21}', \ldots, \boldsymbol{\eta}_{2N}')'.$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

The BLUE of $\boldsymbol{\beta}_1$ and $\boldsymbol{\gamma}$ of (6.5.7) is the GLS estimator.

$$\begin{aligned}\begin{bmatrix}\hat{\boldsymbol{\beta}}_1\\ \hat{\boldsymbol{\gamma}}\end{bmatrix}_{\text{GLS}} &= \left\{\begin{bmatrix}X_1'\\ W'\end{bmatrix}[\Sigma\otimes I_T + \tilde{X}_2\tilde{\Lambda}\tilde{X}_2']^{-1}(X_1, W)\right\}^{-1}\\ &\quad\cdot\left\{\begin{bmatrix}X_1'\\ W'\end{bmatrix}[\Sigma\otimes I_T + \tilde{X}_2\tilde{\Lambda}\tilde{X}_2']^{-1}\mathbf{y}\right\}.\end{aligned} \tag{6.5.8}$$

If Σ is diagonal, the variance–covariance matrix of the stochastic term of (6.5.7) is block-diagonal, with the ith diagonal block equal to

$$\Omega_i = X_{i2}\Lambda X_{i2}' + \sigma_{ii}I_T. \tag{6.5.9}$$

The GLS estimator (6.5.8) can be simplified as

$$\begin{aligned}\begin{bmatrix}\hat{\boldsymbol{\beta}}_1\\ \hat{\boldsymbol{\gamma}}\end{bmatrix}_{\text{GLS}} &= \left[\sum_{i=1}^{N}\begin{bmatrix}X_{i1}'\\ Z_i'X_{i2}'\end{bmatrix}\Omega_i^{-1}(X_{i1}, X_{2i}Z_i)\right]^{-1}\\ &\quad\cdot\left[\sum_{i=1}^{N}\begin{bmatrix}X_{i1}'\\ Z_i'X_{i2}'\end{bmatrix}\Omega_i^{-1}\mathbf{y}_i\right].\end{aligned} \tag{6.5.10}$$

Amemiya (1978b) suggested estimating Λ and σ_{ij} as follows. Let

$$\begin{aligned}\begin{bmatrix}\mathbf{y}_1\\ \vdots\\ \mathbf{y}_N\end{bmatrix} &= \begin{bmatrix}X_{11}\\ \vdots\\ X_{N1}\end{bmatrix}\boldsymbol{\beta}_1 + \begin{bmatrix}X_{12}\\ \mathbf{0}\\ \vdots\\ \mathbf{0}\end{bmatrix}\boldsymbol{\beta}_{21} + \begin{bmatrix}\mathbf{0}\\ X_{22}\\ \vdots\\ \mathbf{0}\end{bmatrix}\boldsymbol{\beta}_{22}\\ &\quad + \cdots + \begin{bmatrix}\mathbf{0}\\ \vdots\\ X_{N2}\end{bmatrix}\boldsymbol{\beta}_{2N} + \begin{bmatrix}\mathbf{u}_1\\ \vdots\\ \mathbf{u}_N\end{bmatrix}.\end{aligned} \tag{6.5.11}$$

Apply the least-squares method to (6.5.11). Denote the resulting estimates by $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_{2i}$, $i = 1, \ldots, N$. Then σ_{ij} can be estimated by

$$\hat{\sigma}_{ij} = \frac{1}{T}(\mathbf{y}_i - X_{i1}\hat{\boldsymbol{\beta}}_1 - \mathbf{X}_{i2}\hat{\boldsymbol{\beta}}_{2i})'(\mathbf{y}_j - \mathbf{X}_{j1}\hat{\boldsymbol{\beta}}_1 - \mathbf{X}_{j2}\hat{\boldsymbol{\beta}}_{2j}), \tag{6.5.12}$$

and $\boldsymbol{\gamma}$ can be estimated by

$$\hat{\boldsymbol{\gamma}} = \left(\sum_{i=1}^{N}Z_i'Z_i\right)^{-1}\left(\sum_{i=1}^{N}Z_i'\hat{\boldsymbol{\beta}}_{2i}\right). \tag{6.5.13}$$

We then estimate Λ by

$$\hat{\Lambda} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\boldsymbol{\beta}}_{2i} - Z_i\hat{\boldsymbol{\gamma}})(\hat{\boldsymbol{\beta}}_{2i} - Z_i\hat{\boldsymbol{\gamma}})'. \tag{6.5.14}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Once consistent estimates of σ_{ij} and Λ are obtained (as both N and T approach infinity), we can substitute them into (6.5.8). The resulting two-stage Aitken estimator of $(\boldsymbol{\beta}_1', \boldsymbol{\gamma}')$ is consistent and asymptotically normally distributed under general conditions. A test of the hypothesis that $\boldsymbol{\gamma} = 0$ can be performed in the usual regression framework using $\hat{\boldsymbol{\gamma}}'_{\text{GLS}}$ Var $(\hat{\boldsymbol{\gamma}}_{\text{GLS}})^{-1}\hat{\boldsymbol{\gamma}}_{\text{GLS}}$, where

$$\text{Var}(\hat{\boldsymbol{\gamma}}_{\text{GLS}}) = [W'\tilde{\Omega}^{-1}W - W'\tilde{\Omega}^{-1}X_1(X_1'\tilde{\Omega}^{-1}X_1)^{-1}X_1'\tilde{\Omega}^{-1}W]^{-1}, \tag{6.5.15}$$

and

$$\tilde{\Omega} = \tilde{X}_2\tilde{\Lambda}\tilde{X}_2' + \Sigma \otimes I_T.$$

6.6 A MIXED FIXED- AND RANDOM-COEFFICIENTS MODEL

6.6.1 Model Formulation

Many of the previously discussed models can be considered as special cases of a general mixed fixed- and random-coefficients model. For ease of exposition, we shall assume that only time-invariant cross-sectional heterogeneity exists.

Suppose that each cross-sectional unit is postulated to be different, so that

$$y_{it} = \sum_{k=1}^{K} \beta_{ki}x_{kit} + \sum_{\ell=1}^{m} \gamma_{\ell i}w_{\ell it} + u_{it}, \quad \begin{matrix} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{matrix} \tag{6.6.1}$$

where $\mathbf{x}_{it}$ and $\mathbf{w}_{it}$ are each $K \times 1$ and $m \times 1$ vector of explanatory variables that are independent of the error of the equation, u_{it}. Stacking the NT observations together, we have

$$\mathbf{y} = X\boldsymbol{\beta} + W\boldsymbol{\gamma} + \mathbf{u}, \tag{6.6.2}$$

where

$$\underset{NT\times NK}{X} = \begin{pmatrix} X_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & X_2 & \ldots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & & & X_N \end{pmatrix},$$

$$\underset{NT\times Nm}{W} = \begin{pmatrix} W_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & W_2 & & \\ \vdots & & \ddots & \\ \mathbf{0} & & & W_N \end{pmatrix},$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

$$\underset{NT\times 1}{\mathbf{u}} = (\mathbf{u}_1', \ldots, \mathbf{u}_N'),$$

$$\underset{NK\times 1}{\boldsymbol{\beta}} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_N')' \quad \text{and} \quad \underset{Nm\times 1}{\boldsymbol{\gamma}} = (\boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_N').$$

Equation (6.6.1), just like (6.6.2), assumes a different behavioral equation relation for each cross-sectional unit. In this situation, the only advantage of pooling is to put the model (6.6.2) into Zellner's (1962) seemingly unrelated regression framework to gain efficiency of the estimates of the individual behavioral equation.

The motivation of a mixed fixed- and random-coefficients model is that though there may be fundamental differences among cross-sectional units, conditioning on these individual specific effects, one may still be able to draw inferences on certain population characteristics through the imposition of a priori constraints on the coefficients of $\mathbf{x}_{it}$ and $\mathbf{w}_{it}$. We assume that there exist two kinds of restrictions, stochastic and fixed restrictions (e.g., Hsiao 1991a, Hsiao et al. 1993) in the form:

A.6.6.1. The coefficients of $\mathbf{x}_{it}$ are assumed to be subject to stochastic restrictions of the form:

$$\boldsymbol{\beta} = A_1\bar{\boldsymbol{\beta}} + \boldsymbol{\alpha}, \tag{6.6.3}$$

where A_1 is an $NK \times L$ matrix with known element, $\bar{\boldsymbol{\beta}}$ is an $L \times 1$ vector of constants, and $\boldsymbol{\alpha}$ is assumed to be (normally distributed) random variables with mean $\mathbf{0}$ and nonsingular constant covariance matrix C and is independent of $\mathbf{x}_{it}$.

A.6.6.2. The coefficients of $\mathbf{w}_{it}$ are assumed to be subject to

$$\boldsymbol{\gamma} = A_2\bar{\boldsymbol{\gamma}}, \tag{6.6.4}$$

where A_2 is an $Nm \times n$ matrix with known elements, and $\bar{\boldsymbol{\gamma}}$ is an $n \times 1$ vector of constants.

Since A_2 is known, we may substitute (6.6.4) into (6.6.2) and write the model as

$$\mathbf{y} = X\boldsymbol{\beta} + \tilde{W}\bar{\boldsymbol{\gamma}} + \mathbf{u} \tag{6.6.5}$$

subject to (6.6.3), where $\tilde{W} = WA_2$.

A.6.6.2 allows for various possible fixed-parameter configurations. For instance, if $\boldsymbol{\gamma}$ is different across cross-sectional units, we can let $A_2 = I_N \otimes I_m$. On the other hand, if we wish to constrain $\boldsymbol{\gamma}_i = \boldsymbol{\gamma}_j$, we can let $A_2 = \mathbf{e}_N \otimes I_m$.

Many of the linear panel data models with unobserved individual specific, but time-invariant heterogeneity can be treated as the special case of the model (6.6.2)–(6.6.4). These include

(1) A common model for all cross-sectional units. If there is no interindividual difference in behavioral patterns, we may let $X = \mathbf{0}$,

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

$A_2 = \mathbf{e}_N \otimes I_m$, so (6.6.2) becomes

$$y_{it} = \mathbf{w}_{it}'\bar{\boldsymbol{\gamma}} + u_{it}. \tag{6.6.6}$$

(2) Different models for different cross-sectional units. When each individual is considered different, then $X = \mathbf{0}$, $A_2 = I_N \otimes I_m$, and (6.6.2) becomes

$$y_{it} = \mathbf{w}_{it}'\boldsymbol{\gamma}_i + u_{it}. \tag{6.6.7}$$

(3) Variable intercept model (e.g., Kuh 1963, or Chapter 3, Section 3.2). If conditional on the observed exogenous variables, the interindividual differences stay constant through time. Let $X = \mathbf{0}$, and

$$A_2 = (I_N \otimes \mathbf{i}_m \vdots \mathbf{e}_N \otimes I^*_{m-1}),\ \bar{\boldsymbol{\gamma}} = (\gamma_{11}, \ldots, \gamma_{N1}, \bar{\gamma}_2, \ldots, \bar{\gamma}_m)',$$

where we arrange $W_i = (\mathbf{e}_T, \mathbf{w}_{i2}, \ldots, \mathbf{w}_{im}), i = 1, \ldots, N$. $\mathbf{i}_m = (1, 0, \ldots, 0)'$,

$$\underset{m\times(m-1)}{I^*_{m-1}} = (\mathbf{0} \vdots I_{m-1})',$$

then (6.6.2) becomes

$$y_{it} = \gamma_{i1} + \bar{\gamma}_2 w_{it2} + \cdots + \bar{\gamma}_m w_{itm} + u_{it}. \tag{6.6.8}$$

(4) Error components model (e.g., Balestra and Nerlove 1966; Wallace and Hussain 1969; or Chapter 3, Section 3.3). When the effects of the individual-specific, time-invariant omitted variables are treated as random variables just like the assumption on the effects of other omitted variables, we can let $X_i = \mathbf{e}_T$, $\boldsymbol{\alpha}' = (\alpha_1, \ldots, \alpha_N)$, $A_1 = \mathbf{e}_N$, $C = \sigma^2_\alpha I_N$, $\bar{\beta}$ be an unknown constant, and $\mathbf{w}_{it}$ not contain an intercept term. Then (6.6.2) becomes

$$y_{it} = \bar{\beta} + \bar{\boldsymbol{\gamma}}'\mathbf{w}_{it} + \alpha_i + u_{it} \tag{6.6.9}$$

(5) Random coefficients model (Swamy 1970, or Chapter 6, Section 6.2.2). Let $Z = \mathbf{0}$, $A_1 = \mathbf{e}_N \otimes I_K$, $C = I_N \otimes \Delta$, we have model (6.2.7).

6.6.2 A Bayes Solution

The formulation of (6.6.5) subject to (6.6.3) can be viewed from a Bayesian perspective as there exist informative prior on $\boldsymbol{\beta}$ (6.6.3), but not on $\bar{\boldsymbol{\gamma}}$. In the classical sampling approach, inferences are made by typically assuming that the probability law generating the observations, $\mathbf{y}$, $f(\mathbf{y}, \boldsymbol{\theta})$, is known, but not the vector of constant parameters $\boldsymbol{\theta}$. Estimators $\hat{\boldsymbol{\theta}}(\mathbf{y})$ of the parameters $\boldsymbol{\theta}$ are chosen as functions of $\mathbf{y}$ so that their sampling distributions, in repeated experiments, are, in some sense, concentrated as closely as possible about the true values of $\boldsymbol{\theta}$. In the Bayesian approach, a different line is taken. First, all quantities, including the parameters, are considered random variables. Second, all probability statements are conditional, so that in making a probability statement

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

it is necessary to refer to the conditioning event as well as the event whose probability is being discussed. Therefore, as part of the model, a prior distribution of the parameter $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$, is introduced. The prior is supposed to express a state of knowledge (or ignorance) about $\boldsymbol{\theta}$ before the data are obtained. Given the probability model $f(y;\boldsymbol{\theta})$, the prior distribution, and the data $\mathbf{y}$, the probability distribution of $\boldsymbol{\theta}$ is revised to $p(\boldsymbol{\theta} \mid \mathbf{y})$, which is called the posterior distribution of $\boldsymbol{\theta}$, according to Bayes' theorem (e.g., Intriligator, Bodkin, and Hsiao 1996).[15]

$$P(\boldsymbol{\theta} \mid \mathbf{y}) \propto P(\boldsymbol{\theta}) f(\mathbf{y} \mid \boldsymbol{\theta}), \tag{6.6.10}$$

where the sign "$\propto$" denoting "is proportional to," with the factor of proportionality being a normalizing constant.

Under the assumption that

A.6.6.3. $\mathbf{u} \sim N(\mathbf{0}, \Omega)$,

we may write the model (6.6.5) as

A.1 Conditional on X, $\tilde{W}$, $\boldsymbol{\beta}$, and $\bar{\boldsymbol{\gamma}}$

$$\mathbf{y} \sim N(X\boldsymbol{\beta} + \tilde{W}\bar{\boldsymbol{\gamma}}, \Omega). \tag{6.6.11}$$

A.2 The prior distributions of $\boldsymbol{\beta}$ and $\bar{\boldsymbol{\gamma}}$ are independent,

$$P(\boldsymbol{\beta}, \bar{\boldsymbol{\gamma}}) = P(\boldsymbol{\beta}) \cdot P(\bar{\boldsymbol{\gamma}}). \tag{6.6.12}$$

A.3 $P(\boldsymbol{\beta}) \sim N(A_1\bar{\boldsymbol{\beta}}, C)$.

A.4 There is no information about $\bar{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{\gamma}}$; therefore $P(\bar{\boldsymbol{\beta}})$ and $P(\bar{\boldsymbol{\gamma}})$ are independent and

$$P(\bar{\boldsymbol{\beta}}) \propto \text{ constant},$$

$$P(\bar{\boldsymbol{\gamma}}) \propto \text{ constant}.$$

Conditional on Ω and C, repeatedly applying the formulas in Appendix 6, yields (Hsiao et al. 1993)

(1) The posterior distribution of $\bar{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{\gamma}}$ given $\mathbf{y}$ is

$$N\left(\begin{pmatrix}\bar{\beta}^* \\ \bar{\boldsymbol{\gamma}}^*\end{pmatrix}, D_1\right), \tag{6.6.13}$$

where

$$D_1 = \left[\begin{pmatrix}A_1'X' \\ \tilde{W}'\end{pmatrix}(\Omega + XCX')^{-1}(XA_1, \tilde{W})\right]^{-1}, \tag{6.6.14}$$

and

$$\begin{pmatrix}\bar{\boldsymbol{\beta}}^* \\ \bar{\boldsymbol{\gamma}}^*\end{pmatrix} = D_1\begin{bmatrix}A_1'X' \\ \tilde{W}'\end{bmatrix}(\Omega + XCX')^{-1}\mathbf{y} \tag{6.6.15}$$

[15] According to Bayes' theorem, the probability of B given A, written as $P(B \mid A)$, equals $P(B \mid A) = \frac{P(A|B)P(B)}{P(A)}$ which is proportional to $P(A \mid B)P(B)$.

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

(2) The posterior distribution of $\boldsymbol{\beta}$ given $\bar{\boldsymbol{\beta}}$ and $\mathbf{y}$ is $N(\boldsymbol{\beta}^*, D_2)$, where

$$D_2 = \{X'[\Omega^{-1} - \Omega^{-1}\tilde{W}(\tilde{W}'\Omega^{-1}\tilde{W})^{-1}\tilde{W}'\Omega^{-1}]X + C^{-1}\}^{-1}, \tag{6.6.16}$$

$$\boldsymbol{\beta}^* = D_2\{X'[\Omega^{-1} - \Omega^{-1}\tilde{W}(\tilde{W}'\Omega^{-1}\tilde{W})^{-1}\tilde{W}'\Omega^{-1}]\mathbf{y} + C^{-1}A_1\bar{\boldsymbol{\beta}}\}. \tag{6.6.17}$$

(3) The (unconditional) posterior distribution of $\boldsymbol{\beta}$ is $N(\boldsymbol{\beta}^{**}, D_3)$, where

$$\begin{aligned} D_3 = \{&X'[\Omega^{-1} - \Omega^{-1}\tilde{W}(\tilde{W}'\Omega^{-1}\tilde{W})^{-1}\tilde{W}'\Omega^{-1}]X + C^{-1} \\ &- C^{-1}A_1(A_1'C^{-1}A_1)^{-1}A_1'C^{-1}\}^{-1}, \end{aligned} \tag{6.6.18}$$

$$\begin{aligned} \boldsymbol{\beta}^{**} &= D_3\{X'[\Omega^{-1} - \Omega^{-1}\tilde{W}(\tilde{W}'\Omega^{-1}\tilde{W})^{-1}\tilde{W}'\Omega^{-1}]\mathbf{y}\} \qquad (6.6.19) \\ &= D_2\{X'[\Omega^{-1} - \Omega^{-1}\tilde{W}(\tilde{W}'\Omega^{-1}\tilde{W})^{-1}\tilde{W}'\Omega^{-1}]X\hat{\boldsymbol{\beta}} + C^{-1}A_1\bar{\boldsymbol{\beta}}^*\}, \end{aligned}$$

where $\hat{\boldsymbol{\beta}}$ is the GLS estimate of (6.6.5),

$$\begin{aligned} \hat{\boldsymbol{\beta}} = &\{X'[\Omega^{-1} - \Omega^{-1}\tilde{W}(\tilde{W}'\Omega^{-1}\tilde{W})^{-1}\tilde{W}'\Omega^{-1}]X\}^{-1} \\ &\cdot \{X'[\Omega^{-1} - \Omega^{-1}\tilde{W}(\tilde{W}'\Omega^{-1}\tilde{W})^{-1}\tilde{W}'\Omega^{-1}]\mathbf{y}\}. \end{aligned} \tag{6.6.20}$$

Given a quadratic loss function of the error of the estimation, a Bayes point estimate is the posterior mean. The posterior mean of $\bar{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{\gamma}}$ (6.6.15) is the GLS estimator of the model (6.6.5) after substituting the restriction (6.6.3),

$$\mathbf{y} = XA_1\bar{\boldsymbol{\beta}} + \tilde{W}\bar{\boldsymbol{\gamma}} + \mathbf{v}, \tag{6.6.21}$$

where $\mathbf{v} = X\boldsymbol{\alpha} + \mathbf{u}$. However, the posterior mean of $\boldsymbol{\beta}$ is not the GLS estimator of (6.6.5). It is the weighted average between the GLS estimator of $\boldsymbol{\beta}$ and the overall mean $\bar{\boldsymbol{\beta}}$ (6.6.17) or $\bar{\boldsymbol{\beta}}^*$ (6.6.19), with the weights proportional to the inverse of the precision of respective estimates. The reason is that although both (6.6.2) and (6.6.5) allow the coefficients to be different across cross-sectional units, (6.6.3) has imposed additional prior information that $\boldsymbol{\beta}$ are randomly distributed with mean $A_1\bar{\boldsymbol{\beta}}$. For (6.6.2), the best linear predictor for an individual outcome is to substitute the best linear unbiased estimator of the individual coefficients into the individual equation. For model of (6.6.5) and (6.6.3), because the expected $\boldsymbol{\beta}_i$ is the same across i and the actual difference can be attributed to a chance outcome, additional information about $\boldsymbol{\beta}_i$ may be obtained by examining the behavior of others, hence (6.6.17) or (6.6.19).

In the special case of error components model (6.6.9), $X = I_N \otimes \mathbf{e}_T$. Under the assumption that $\mathbf{w}_{it}$ contains an intercept term (i.e., $\bar{\beta} = 0$) and u_{it} is i.i.d., the Bayes estimator ((6.6.15)) of $\bar{\boldsymbol{\gamma}}$ is simply the GLS estimator of (6.6.21), $\bar{\boldsymbol{\gamma}}^*$. The Bayes estimator of α_i ((6.6.17)) is

$$\alpha_i^{**} = \left(\frac{T\sigma_\alpha^2}{T\sigma_\alpha^2 + \sigma_u^2}\right)\hat{\bar{v}}_i, \tag{6.6.22}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

where $\hat{\bar{v}}_i = \frac{1}{T}\sum_{\tau=1}^{T}\hat{v}_{it}$ and $\hat{v}_{it} = y_{it} - \bar{\boldsymbol{\gamma}}^{*}\mathbf{w}_{it}$. Substituting $\bar{\boldsymbol{\gamma}}^{*}$, and α_i^{**} for the unknown $\bar{\boldsymbol{\gamma}}$, and α_i in (6.6.9), Wansbeek and Kapteyn (1978) and Taub (1979) show that

$$\hat{y}_{i,T+S} = \bar{\boldsymbol{\gamma}}^{*'}\mathbf{w}_{i,t+s} + \alpha_i^{**} \tag{6.6.23}$$

is the best linear predictor (BLUP) for the ith individual s periods ahead.[16]

6.6.3 Random or Fixed Differences?

6.6.3.1 An Example of the Contrast between Individual and Pooled Parameter Estimates

In a classical framework, it makes no sense to predict the independently drawn random variable $\boldsymbol{\beta}_i$ (or $\boldsymbol{\alpha}_i$). However, in panel data, we actually operate with two dimensions – a cross-sectional dimension and a time series dimension. Even though $\boldsymbol{\beta}_i$ is an independently distributed random variable across i, once a particular $\boldsymbol{\beta}_i$ is drawn, it stays constant over time. Therefore, it makes sense to predict $\boldsymbol{\beta}_i$. The classical predictor of $\boldsymbol{\beta}_i$ is the GLS estimator of the model (6.6.5). The Bayes predictor (6.6.19) is the weighted average between the GLS estimator of $\boldsymbol{\beta}$ for the model (6.6.5) and the overall mean $A_1\bar{\boldsymbol{\beta}}$ if $\bar{\boldsymbol{\beta}}$ is known or $A_1\bar{\boldsymbol{\beta}}^{*}$ if $\bar{\boldsymbol{\beta}}$ is unknown with the weights proportional to the inverse of the precisions of respective estimates. The Bayes estimator of the individual coefficients, $\boldsymbol{\beta}_i$, "shrinks" the GLS estimator of $\boldsymbol{\beta}_i$ toward the grand mean $\bar{\boldsymbol{\beta}}$ or $\bar{\boldsymbol{\beta}}^{*}$. The reason for doing so stems from de Finneti's (1964) exchangeability assumption. When there are not enough time series observations to allow for precise estimation of individual $\boldsymbol{\beta}_i$ (i.e., T is small), additional information about $\boldsymbol{\beta}_i$ may be obtained by examining the behavior of others because the expected response is assumed the same and the actual differences in response among individuals are the work of a chance mechanism.

Table 6.1 presents the Canadian route specific estimates of the demand for customer-dialed long distance service over 920 miles (long-haul) based on quarterly data from 1980.I to 1989.IV (Hsiao, Appelbe, and Dineen 1993). Some of the point-to-point individual route estimates (unconstrained model) of the price and income coefficients have the wrong signs (Table 6.1, column 2), perhaps because of multicollinearity. However, when one invokes the representative consumer argument by assuming that consumers respond in more or less the same way to price and income changes, thus assuming the coefficients of these variables across routes are considered random draws from a common population with constant mean and variance–covariance matrix, but also allows the route-specific effects to exist by assuming that the coefficients of the intercept and seasonal dummies are fixed and different for different routes, all the

[16] When u_{it} is serially correlated, see Baltagi and Li (1992). For the asymptotic mean square error when the coefficients and error components parameters are estimated, see Baillie and Baltagi (1999).

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Table 6.1. *Long-haul regression coefficients*[a]

Route	Price coefficient, unconstrained	Mixed coefficients
1	−0.0712(−0.15)	−0.2875(N/A)
2	0.1694(0.44)	−0.0220(N/A)
3	−1.0142(−5.22)	−0.7743(N/A)
4	−0.4874(−2.29)	−0.1686(N/A)
5	−0.3190(−2.71)	−0.2925(N/A)
6	0.0365(0.20)	−0.0568(N/A)
7	−0.3996(−3.92)	−0.3881(N/A)
8	−0.1033(−0.95)	−0.2504(N/A)
9	−0.3965(−4.22)	−0.2821(N/A)
10	−0.6187(−4.82)	−0.5934(N/A)
Average	N/A	−0.3116
Route	Income coefficient	
1	1.4301(3.07)	0.4740(N/A)
2	−0.348(−0.09)	0.2679(N/A)
3	0.3698(1.95)	0.3394(N/A)
4	0.2497(0.70)	0.3145(N/A)
5	0.5556(2.71)	0.3501(N/A)
6	0.1119(0.95)	0.1344(N/A)
7	0.9197(8.10)	0.5342(N/A)
8	0.3886(3.88)	0.5255(N/A)
9	0.6688(6.16)	0.5648(N/A)
10	0.1928(2.39)	0.2574(N/A)
Average	N/A	0.3762

[a] t-statistics in parentheses.
Source: Hsiao, Appelbe, and Dineen (1993, Table 3).

estimated route specific price and income coefficients have the correct signs (Table 6.1, column 3).

6.6.3.2 An Example of Prediction Comparison

When homogeneity is rejected by the data, whether to treat unobserved heterogeneity as fixed or random has paramount importance in panel data modeling. For instance, in a study of Ontario, Canada regional electricity demand, Hsiao et al. (1989) estimate a model of the form

$$y_{it} = \gamma_i y_{i,t-1} + \boldsymbol{\delta}_i' \mathbf{d}_{it} + \boldsymbol{\beta}_i' \mathbf{x}_{it} + u_{it}, \tag{6.6.24}$$

where y_{it} denotes the logarithm of monthly kilowatt-hour or kilowatt demand for region i at time t; $\mathbf{d}_{it}$ denotes 12 monthly dummies; and $\mathbf{x}_{it}$ denotes climatic

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Table 6.2. *Root-mean-square prediction error of log kilowatt-hours (one-period-ahead forecast)*

Municipality	Root Mean Square Error			
	Region-specific	Pooled	Random coefficients	Mixed
Hamilton	0.0865	0.0535	0.0825	0.0830
Kitchener–Waterloo	0.0406	0.0382	0.0409	0.0395
London	0.0466	0.0494	0.0467	0.0464
Ottawa	0.0697	0.0523	0.0669	0.0680
St. Catharines	0.0796	0.0724	0.0680	0.0802
Sudbury	0.0454	0.0857	0.0454	0.0460
Thunder Bay	0.0468	0.0615	0.0477	0.0473
Toronto	0.0362	0.0497	0.0631	0.0359
Windsor	0.0506	0.0650	0.0501	0.0438
Unweighted average	0.0558	0.0586	0.0568	0.0545
Weighted average[a]	0.0499	0.0525	0.0628	0.0487

[a] The weight is kilowatt-hours of demand in the municipality in June 1985.
Source: Hsiao et al. (1989, p. 584).

factor and the logarithm of income, own price, and price of its close substitutes, all measured in real terms. Four different specifications are considered:

1. The coefficients $\boldsymbol{\theta}_i' = (\gamma_i, \boldsymbol{\delta}_i', \boldsymbol{\beta}_i')$ are fixed and different for different region.
2. The coefficients $\boldsymbol{\theta}_i = \boldsymbol{\theta}' = (\gamma, \boldsymbol{\delta}', \boldsymbol{\beta}')$ for all i.
3. The coefficient vectors $\boldsymbol{\theta}_i$ are randomly distributed with common mean $\boldsymbol{\theta}$ and covariance matrix Δ.
4. The coefficients $\boldsymbol{\beta}_i$ are randomly distributed with common mean $\bar{\boldsymbol{\beta}}$ and covariance matrix Δ_{11}, and the coefficients γ_i and $\boldsymbol{\delta}_i$ are fixed and different for different i.

Monthly data for Hamilton, Kitchener-Waterloo, London, Ottawa, St. Catherines, Sudbury, Thunder Bay, Toronto, and Windsor from January 1967 to December 1982 are used to estimate these four different specifications. Comparisons of the one-period ahead root mean square prediction error

$$\sqrt{\sum_{t=T+1}^{T+f} (y_{it} - \hat{y}_{it})^2 / f}$$

from January 1983 to December 1986 are summarized in Tables 6.2 and 6.3. As one can see from these tables, the simple pooling (model 2) and random-coefficients (model 3) formulations on average yield less precise prediction for regional demand. The mixed fixed- and random-coefficients model (model 4) performs the best. It is interesting to note that combining information across

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Table 6.3. *Root-mean-square prediction error of log kilowatts (one-period-ahead forecast)*

	Root Mean Square Error			
Municipality	Regional specific	Pooled	Random coefficients	Mixed
Hamilton	0.0783	0.0474	0.0893	0.0768
Kitchener–Waterloo	0.0873	0.0440	0.0843	0.0803
London	0.0588	0.0747	0.0639	0.0586
Ottawa	0.0824	0.0648	0.0846	0.0768
St. Catharines	0.0531	0.0547	0.0511	0.0534
Sudbury	0.0607	0.0943	0.0608	0.0614
Thunder Bay	0.0524	0.0597	0.0521	0.0530
Toronto	0.0429	0.0628	0.0609	0.0421
Windsor	0.0550	0.0868	0.0595	0.0543
Unweighted average	0.0634	0.0655	0.0674	0.0619
Weighted average[a]	0.0558	0.0623	0.0673	0.0540

[a] The weight is kilowatt-hours of demand in the municipality in June 1985.
Source: Hsiao et al. (1989, p. 584).

regions together with a proper account of regional-specific factors is capable of yielding better predictions for regional demand than the approach of simply using regional-specific data (model 1).

6.6.3.3 Model Selection

The preceding example demonstrates that the way in which individual heterogeneity is taken into account makes a difference in the accuracy of inference. The various estimation methods discussed so far presuppose that we know which coefficients should be treated as fixed (and different) and which coefficients should be treated as random. In practice, we have very little prior information on selecting the appropriate specifications. Various statistical tests have been suggested to select an appropriate formulation (e.g., Breusch and Pagan 1979; Hausman 1978 or Chapter 6, Section 6.2.2.4). However, all these tests essentially exploit the implication of certain formulation in a specific framework. They are indirect in nature. The distribution of a test statistic is derived under a specific null, but the alternative is composite. The rejection of a null does not automatically imply the acceptance of a specific alternative. It would appear more appropriate to treat the fixed coefficients, random coefficients, or various forms of mixed fixed- and random-coefficients models as different models and use model selection criteria to select an appropriate specification (Hsiao and Sun 2000). For instance, well known model selection criterion such as Akaike (1973) information criteria or Schwarz (1978) Bayesian information criteria that selects the model H_j among $j = 1, \ldots, J$ different specifications

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

if it yields the smallest value of

$$-2\log f(\mathbf{y} \mid H_j) + 2m_j, \quad j = 1\ldots., J, \tag{6.6.25}$$

or

$$-2\log f(\mathbf{y} \mid H_j) + m_j \log NT, \quad j = 1\ldots., J, \tag{6.6.26}$$

can be used, where $\log f(\mathbf{y} \mid H_j)$ and m_j denote the log-likelihood values of $\mathbf{y}$ and the number of unknown parameters of model H_j. Alternatively, Hsiao (1995) and Min and Zellner (1993) suggest selecting the model that yields the highest predictive density. In this framework, time series observations are divided into two periods, 1 to T_1, denoted by $\mathbf{y}^1$, and $T_1 + 1$ to T, denoted by $\mathbf{y}^2$. The first T_1 observations are used to obtain the probability distribution of the parameters associated with H_j, say $\boldsymbol{\theta}^j$, $P(\boldsymbol{\theta}^j \mid \mathbf{y}^1)$. The predictive density is then evaluated as

$$\int f(\mathbf{y}^2 \mid \boldsymbol{\theta}^j) p(\boldsymbol{\theta}^j \mid \mathbf{y}^1) d\boldsymbol{\theta}^j, \tag{6.6.27}$$

where $f(\mathbf{y}^2 \mid \boldsymbol{\theta}^j)$ is the density of $\mathbf{y}^2$ conditional on $\boldsymbol{\theta}^j$. Given the sensitivity of Bayesian approach to the choice of prior distribution the advantage of using (6.6.27) is that the choice of a model does not have to depend on the prior. One can use the noninformative (or diffuse) prior to derive $P(\boldsymbol{\theta}^j \mid \mathbf{y}^1)$. It is also consistent with the theme that "a severe test for an economic theory, the only test and the ultimate test is its ability to predict" (Klein 1988; p. 21; see also Friedman 1953).

When $\mathbf{y}^2$ contains only a limited number of observations, the choice of model in terms of predictive density may become heavily sample dependent. If too many observations are put in $\mathbf{y}^2$, then a great deal of sample information is not utilized to estimate unknown parameters. One compromise is to modify (6.6.27) by recursively updating the estimates,

$$\begin{aligned} &\int f(\mathbf{y}_T \mid \boldsymbol{\theta}^j, \mathbf{y}^{T-1}) P(\boldsymbol{\theta}^j \mid \mathbf{y}^{T-1}) d\boldsymbol{\theta}^j \\ &\cdot \int f(\mathbf{y}_{T-1} \mid \boldsymbol{\theta}^j, \mathbf{y}^{T-2}) P(\boldsymbol{\theta}^j \mid \mathbf{y}^{T-2}) d\boldsymbol{\theta}^j \\ &\ldots \int f(\mathbf{y}_{T_1+1} \mid \boldsymbol{\theta}^j, \mathbf{y}^1) P(\boldsymbol{\theta}^j \mid \mathbf{y}^1) d\boldsymbol{\theta}^j, \end{aligned} \tag{6.6.28}$$

where $P(\boldsymbol{\theta}^j \mid y^T)$ denotes the posterior distribution of $\boldsymbol{\theta}$ given observations from 1 to T. While the formula may look formidable, it turns out that the Bayes updating formula is fairly straightforward to compute. For instance, consider the model (6.6.5). Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \bar{\boldsymbol{\gamma}})$ and $\boldsymbol{\theta}_t$ and V_t denote the posterior mean and variance of $\boldsymbol{\theta}$ based on the first t-observations; then

$$\boldsymbol{\theta}_t = V_{t-1}(Q_t' \Omega^{-1} \mathbf{y}_t + V_{t-1}^{-1} \boldsymbol{\theta}_{t-1}), \tag{6.6.29}$$

$$V_t = (Q_t' \Omega^{-1} Q_t + V_{t-1}^{-1})^{-1}, \quad t = T_1 + 1, \ldots, T, \tag{6.6.30}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

and

$$P(\mathbf{y}_{t+1} \mid \mathbf{y}^t) = \int P(\mathbf{y}_{t+1} \mid \theta, \mathbf{y}^t) P(\boldsymbol{\theta} \mid \mathbf{y}^t) d\boldsymbol{\theta}$$
$$\sim N(Q_{t+1}\boldsymbol{\theta}_t, \Omega + Q_{t+1} V_t Q'_{t+1}), \tag{6.6.31}$$

where $\mathbf{y}'_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})$, $Q_t = (\mathbf{x}'_t, \mathbf{w}'_t)$, $\mathbf{x}_t = (\mathbf{x}_{1t}, \ldots, \mathbf{x}_{Nt})$, $\mathbf{w}_t = (\mathbf{w}_{1t}, \ldots, \mathbf{w}_{Nt})$, $\Omega = E\mathbf{u}_t\mathbf{u}'_t$, and $\mathbf{u}'_t = (u_{1t}, \ldots, u_{Nt})$ (Hsiao et al. 1993).

Hsiao and Sun (2000) have conducted limited Monte Carlo studies to evaluate the performance of these model selection criteria in selecting the random, fixed, and mixed random–fixed coefficients specification. They all appear to have a very high percentage in selecting the correct specification.

6.7 DYNAMIC RANDOM-COEFFICIENTS MODELS

For ease of exposition and without loss of the essentials, instead of considering generalizing (6.6.5) into the dynamic model, in this section we consider the generalization of random coefficients model (6.2.1) to the dynamic model of the form[17]

$$y_{it} = \gamma_i y_{i,t-1} + \boldsymbol{\beta}'_i \mathbf{x}_{it} + u_{it}, \quad |\gamma_i| < 1, \quad \begin{matrix} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{matrix} \tag{6.7.1}$$

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 (i.i.d.) over t with mean 0 and variance $\sigma^2_{u_i}$ and is independent across i. The coefficients $\boldsymbol{\theta}_i = (\gamma_i, \boldsymbol{\beta}'_i)'$ are assumed to be independently distributed across i with mean $\bar{\boldsymbol{\theta}} = (\bar{\gamma}, \bar{\boldsymbol{\beta}}')'$ and covariance matrix Δ. Let

$$\boldsymbol{\theta}_i = \bar{\boldsymbol{\theta}} + \boldsymbol{\alpha}_i, \tag{6.7.2}$$

where $\boldsymbol{\alpha}_i = (\alpha_{i1}, \boldsymbol{\alpha}'_{i2})$; we have

$$E\boldsymbol{\alpha}_i = \mathbf{0}, \quad E\boldsymbol{\alpha}_i\boldsymbol{\alpha}'_j = \Delta \text{ if } i = j \text{ and } \mathbf{0} \text{ otherwise}, \tag{6.7.3}$$

[17] We are concerned only with the estimation of the short-run adjustment coefficient γ. For a discussion of estimating the long-run coefficient, see Pesaran and Smith (1995); Pesaran and Zhao (1999); Pesaran, Shin, and Smith (1999); and Phillips and Moon (1999, 2000).

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

and[18]

$$E\boldsymbol{\alpha}_i \mathbf{x}'_{jt} = \mathbf{0}. \tag{6.7.4}$$

Stacking the T time series observations of the ith individuals in matrix form yields

$$\underset{T\times 1}{\mathbf{y}_i} = Q_i\boldsymbol{\theta}_i + \mathbf{u}_i, \quad i = 1, \ldots, N. \tag{6.7.5}$$

where $\mathbf{y}_i = (y_{i1}\ldots., y_{iT})'$, $Q_i = (\mathbf{y}_{i,-1}, X_i)$, $\mathbf{y}_{i,-1} = (y_{i0}, \ldots, y_{i,T-1})'$, $X_i = (\mathbf{x}_{i1}\ldots., \mathbf{x}_{iT})'$, $\mathbf{u}_i = (u_{i1}\ldots., u_{iT})'$, and for ease of exposition, we assume that y_{i0} are observable.[19]

We note that because $y_{i,t-1}$ despends on γ_i, $EQ_i\boldsymbol{\alpha}'_i \neq \mathbf{0}$, that is, the independence between the explanatory variables and $\boldsymbol{\alpha}_i$ (6.2.6) is violated. Substituting $\boldsymbol{\theta}_i = \bar{\boldsymbol{\theta}} + \boldsymbol{\alpha}_i$ into (6.7.5) yields

$$\mathbf{y}_i = Q_i\bar{\boldsymbol{\theta}} + \mathbf{v}_i, \quad i = 1\ldots., N, \tag{6.7.6}$$

where

$$\mathbf{v}_i = Q_i\boldsymbol{\alpha}_i + \mathbf{u}_i. \tag{6.7.7}$$

[18] The strict exogeneity condition (6.7.4) of $\mathbf{x}_{it}$ is crucial in the identification of dynamic random-coefficients model. Chamberlain (1993) has given an example of the lack of identification of γ in a model of the form

$$y_{it} = \gamma y_{i,t-1} + \boldsymbol{\beta}_i x_{it} + \alpha_i + u_{it},$$

where x_{it} takes either 0 or 1. Because $E(\alpha_i \mid \mathbf{x}_i, \mathbf{y}_{i,-1})$ is unrestricted, the only moments that are relevant for the identification of γ are

$$E(\Delta y_{it} - \gamma\Delta y_{i,t-1} \mid \mathbf{x}_i^{t-1}, \mathbf{y}_i^{t-2}) = E(\beta_i \Delta x_{it} \mid \mathbf{x}_i^{t-1}, \mathbf{y}_i^{t-2}), \quad t = 2, \ldots, T,$$

where $\mathbf{x}_i^t = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{it})$, $\mathbf{y}_i^t = (y_{i0}, \ldots, y_{it})$. Let $\mathbf{w}_i^t = (\mathbf{x}_i^t, \mathbf{y}_i^t)$, the above expression is equivalent to the following two conditions:

$$\begin{aligned} &D(\Delta y_{it} - \gamma\Delta y_{i,t-1} \mid \mathbf{w}_i^{t-2}, \mathbf{x}_{i,t-1} = 0) \\ &\quad = E(\beta_i \mid \mathbf{w}_i^{t-2}, \mathbf{x}_{i,t-1} = 0) P_r(x_{it} = 1 \mid \mathbf{x}_i^{t-2}, x_{i,t-1} = 0), \end{aligned}$$

and

$$\begin{aligned} &E(\Delta y_{it} - \gamma\Delta y_{i,t-1} \mid \mathbf{x}_i^{t-2}, x_{i,t-1} = 1) \\ &\quad = -E(\boldsymbol{\beta}_i \mid \mathbf{w}_i^{t-2}, x_{i,t-1} = 1) P_r(x_{it} = 0 \mid \mathbf{w}_i^{t-2}, x_{i,t-1} = 1) \end{aligned}$$

If $E(\beta_i \mid \mathbf{w}_i^{t-2}, x_{i,t-1} = 0)$ and $E(\beta_i \mid \mathbf{w}_i^{t-2}, x_{i,t-1} = 1)$ are unrestricted and T is fixed, the autoregressive parameter γ cannot be identified from the above two equations.

[19] We assume that $T(>3)$ is large enough to identify γ and β. For an example of lack of identification when $T = 3$ and y_{it} is binary, see Chamberlain (1993) or Arellano and Honoré (2001); see also Chapter 7.

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Since

$$y_{i,t-1} = \sum_{j=0}^{\infty} (\bar{\gamma} + \alpha_{i1})^j \mathbf{x}'_{i,t-j-1}(\bar{\boldsymbol{\beta}} + \boldsymbol{\alpha}_{i2}) + \sum_{j=0}^{\infty} (\bar{\gamma} + \alpha_{i1})^j u_{i,t-j-1}, \tag{6.7.8}$$

it follows that $E(\mathbf{v}_i \mid Q_i) \neq \mathbf{0}$. Therefore, contrary to the static case, the least-squares estimator of the common mean, $\bar{\boldsymbol{\theta}}$ is inconsistent.

Equations (6.7.7) and (6.7.8) also demonstrate that the covariance matrix of $\mathbf{v}_i$, V, is not easily derivable. Thus, the procedure of premultiplying (6.7.6) by $V^{-1/2}$ to transform the model into the one with serially uncorrelated error is not implementable. Neither does the instrumental variable method appear implementable because the instruments that are uncorrelated with $\mathbf{v}_i$ are most likely uncorrelated with Q_i as well.

Pesaran and Smith (1995) have noted that as $T \to \infty$, the least-squares regression of $\mathbf{y}_i$ on Q_i 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,

$$\hat{\bar{\boldsymbol{\theta}}} = \frac{1}{N} \sum_{i=1}^{N} \hat{\boldsymbol{\theta}}_i. \tag{6.7.9}$$

The mean group estimator (6.7.9) is consistent and asymptotically normally distributed so long as $\sqrt{N}/T \to 0$ as both N and $T \to \infty$ (Hsiao, Pesaran, and Tahmiscioglu 1999).

However, panels with large T are typically the exception in economics. Nevertheless, under the assumption that y_{i0} are fixed and known and $\boldsymbol{\alpha}_i$ and u_{it} are independently normally distributed, we can implement the Bayes estimator of $\bar{\boldsymbol{\theta}}$ conditional on σ_i^2 and Δ using the formula (6.6.13) just as in the mixed model case discussed in Section 6.6. The Bayes estimator condition on Δ and σ_i^2 is equal to

$$\hat{\bar{\boldsymbol{\theta}}}_B = \left\{ \sum_{i=1}^{N} \left[\sigma_i^2 (Q_i' Q_i)^{-1} + \Delta \right]^{-1} \right\}^{-1} \sum_{i=1}^{N} \left[\sigma_i^2 (Q_i' Q_i)^{-1} + \Delta \right]^{-1} \hat{\boldsymbol{\theta}}_i, \tag{6.7.10}$$

which is a weighted average of the least-squares estimator of individual units, with the weights being inversely proportional to individual variances. When $T \to \infty$, $N \to \infty$ and $\sqrt{N}/T^{3/2} \to 0$, the Bayes estimator is asymptotically equivalent to the mean group estimator (6.7.9).

In practice, the variance components, σ_i^2 and Δ, are rarely known, so the Bayes estimator (6.7.10) is rarely feasible. One approach is to substitute the consistently estimated σ_i^2 and Δ, say (6.2.11) and (6.2.12), into the formula (6.7.10), and treat them as if they were known. For ease of reference, we shall

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

call (6.7.10) with known σ_i^2 and Δ the infeasible Bayes estimator. We shall call the estimator obtained by substituting σ_i^2 and Δ in (6.7.10) by their consistent estimates, say (6.2.11) and (6.2.12), the empirical Bayes estimator.

The other approach is to follow Lindley and Smith (1972) by assuming that the prior distributions of σ_i^2 and Δ are independent and are distributed as

$$P\left(\Delta^{-1}, \sigma_1^2, \ldots, \sigma_N^2\right) = W(\Delta^{-1} \mid (\rho R)^{-1}, \rho) \prod_{i=1}^{N} \sigma_i^{-1}, \tag{6.7.11}$$

where W represents the Wishart distribution with scale matrix (ρR) and degrees of freedom ρ (e.g., Anderson 1985). Incorporating this prior into the model (6.7.1)–(6.7.2), we can obtain the marginal posterior densities of the parameters of interest by integrating out σ_i^2 and Δ from the joint posterior density. However, the required integrations do not yield closed form solutions. Hsiao, Pesaran, and Tahmiscioglu (1999) have suggested using Gibbs sampler to calculate marginal densities.

The Gibbs sampler is an iterative Markov Chain Monte Carlo method that requires only the knowledge of the full conditional densities of the parameter vector (e.g., Gelfand and Smith 1990). Starting from some arbitrary initial values, say $(\boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)}, \ldots, \boldsymbol{\theta}_k^{(0)})$ for a parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1 \ldots, \boldsymbol{\theta}_k)$, it samples alternatively from the conditional density of each component of the parameter vector conditional on the values of other components sampled in the latest iteration. That is:

(1) Sample $\boldsymbol{\theta}_1^{(j+1)}$ from $P(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2^{(j)}, \boldsymbol{\theta}_3^{(j)}, \ldots, \boldsymbol{\theta}_k^{(j)}, \mathbf{y})$
(2) Sample $\boldsymbol{\theta}_2^{(j+1)}$ from $P(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1^{(j+1)}, \boldsymbol{\theta}_3^{(j)}, \ldots, \boldsymbol{\theta}_k^{(j)}, \mathbf{y})$
$\vdots$
(k) Sample $\boldsymbol{\theta}_k^{(j+1)}$ from $P(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_1^{(j+1)}, \ldots, \boldsymbol{\theta}_{k-1}^{(j+1)}, \mathbf{y})$

The vectors $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(k)}$ form a Markov Chain, with transition probability from stage $\boldsymbol{\theta}^{(j)}$ to the next stage $\boldsymbol{\theta}^{(j+1)}$ being

$$K(\boldsymbol{\theta}^{(j)}, \boldsymbol{\theta}^{(j+1)}) = P\left(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2^{(j)}, \ldots, \boldsymbol{\theta}_k^{(j)}, \mathbf{y}\right) P\left(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1^{(j+1)}, \boldsymbol{\theta}_3^{(j)}, \ldots, \boldsymbol{\theta}_k^{(j)}, y\right)$$
$$\ldots P\left(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_1^{(j+1)}, \ldots, \boldsymbol{\theta}_{k-1}^{(j+1)}, \mathbf{y}\right).$$

As the number of iterations j approaches infinity, the sampled values in effect can be regarded as drawing from true joint and marginal posterior densities. Moreover, the ergodic averages of functions of the sample values will be consistent estimations of their expected values.

Under the assumption that the prior of $\bar{\theta}$ is $N(\bar{\boldsymbol{\theta}}^*, \Psi)$, the relevant conditional distributions that are needed to implement the Gibbs sampler for (6.7.1)–(6.7.2)

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

are easily obtained from

$$P\left(\boldsymbol{\theta}_i \mid \mathbf{y}, \bar{\boldsymbol{\theta}}, \Delta^{-1}, \sigma_1^2, \ldots, \sigma_N^2\right)$$
$$\sim N\left\{A_i(\sigma_i^{-2} Q_i' \mathbf{y}_i + \Delta^{-1}\bar{\theta}), A_i\right\}, \quad i = 1, \ldots, N,$$
$$P(\bar{\boldsymbol{\theta}} \mid \mathbf{y}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, \Delta^{-1}, \sigma_1^2, \ldots, \sigma_N^2) \sim N\left\{D(N\Delta^{-1}\hat{\bar{\boldsymbol{\theta}}} + \Psi^{-1}\boldsymbol{\theta}^*), B\right\}$$
$$P\left(\Delta^{-1} \mid \mathbf{y}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, \bar{\boldsymbol{\theta}}, \sigma_1^2, \ldots, \sigma_N^2\right)$$
$$\sim W\left[\left(\sum_{i=1}^{N}(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})' + \rho R\right)^{-1}, \rho + N\right],$$
$$P\left(\sigma_i^2 \mid \mathbf{y}_i, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, \bar{\boldsymbol{\theta}}, \Delta^{-1}\right)$$
$$\sim IG[T/2, (\mathbf{y}_i - Q_i\boldsymbol{\theta}_i)'(\mathbf{y}_i - Q_i\boldsymbol{\theta}_i)/2], i = 1\ldots., N,$$

where $A_i = (\sigma_i^{-2} Q_i' Q_i + \Delta^{-1})^{-1}$, $D = (N\Delta^{-1} + \Psi^{-1})^{-1}$, $\hat{\bar{\boldsymbol{\theta}}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\theta}_i$, and IG denotes the inverse gamma distribution.

Hsiao, Pesaran, and Tahmiscioglu (1999) have conducted Monte Carlo experiments to study the finite sample properties of (6.7.10), referred as infeasible Bayes estimator; the Bayes estimator obtained through the Gibbs sampler, referred as hierarchical Bayes estimator; the empirical Bayes estimator; the group mean estimator (6.7.8); the bias corrected group mean estimator obtained by directly correcting the finite T bias of the least squares estimator, $\hat{\boldsymbol{\theta}}_i$, using the formula of Kiviet (1995); Kiviet and Phillips (1993); and then taking the average; and the pooled least-squares estimator. Table 6.4 presents the bias of the different estimators of $\bar{\gamma}$ for $N = 50$ and $T = 5$ or 20. The infeasible Bayes estimator performs very well. It has small bias even for $T = 5$. For $T = 5$, its bias falls within the range of 3 to 17 percent. For $T = 20$, the bias is at most about 2 percent. The hierarchical Bayes estimator also performs well,[20] followed by the empirical Bayes estimator when T is small but improves quickly as T increases. The empirical Bayes estimator gives very good results even for $T = 5$ in some cases but the bias also appears to be quite high in certain other cases. As T gets larger its bias decreases considerably. The mean group and the bias corrected mean group estimator both have large bias when T is small, with the bias-corrected mean group estimator performing slightly better. However, the performance of both improve as T increases, and both are still much better than the least-squares estimator. The least-squares estimator yields significant bias and its bias persists as T increases.

The Bayes estimator is derived under the assumption that the initial observations, y_{i0}, are fixed constants. As discussed in Chapter 4 or Anderson and Hsiao (1981, 1982), this assumption is clearly unjustifiable for a panel with

[20] The $\Psi^{-1} = 0$, $\rho = 2$ and R equal to the Swamy estimate of Δ are used to implement the hierarchical Bayes estimator.

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Table 6.4. *Bias of the short-run coefficient* $\bar{\gamma}$

			Bias					
T	$\bar{\gamma}$		Pooled OLS	Mean group	Bias-corrected mean group	Infeasible Bayes	Empirical Bayes	Hierarchical Bayes
5	1	0.3	0.36859	−0.23613	−0.14068	0.05120	−0.12054	−0.02500
	2	0.3	0.41116	−0.23564	−0.14007	0.04740	−0.11151	−0.01500
	3	0.6	1.28029	−0.17924	−0.10969	0.05751	−0.02874	0.02884
	4	0.6	1.29490	−0.18339	−0.10830	0.06879	−0.00704	0.06465
	5	0.3	0.06347	−0.26087	−0.15550	0.01016	−0.18724	−0.10068
	6	0.3	0.08352	−0.26039	−0.15486	0.01141	−0.18073	−0.09544
	7	0.6	0.54756	−0.28781	−0.17283	0.05441	−0.12731	−0.02997
	8	0.6	0.57606	−0.28198	−0.16935	0.06258	−0.10366	−0.01012
20	9	0.3	0.44268	−0.07174	−0.01365	0.00340	−0.00238	0.00621
	10	0.3	0.49006	−0.06910	−0.01230	0.00498	−0.00106	0.00694
	11	0.35	0.25755	−0.06847	−0.01209	−0.00172	−0.01004	−0.00011
	12	0.35	0.25869	−0.06644	−0.01189	−0.00229	−0.00842	0.00116
	13	0.3	0.07199	−0.07966	−0.01508	−0.00054	−0.01637	−0.00494
	14	0.3	0.09342	−0.07659	−0.01282	0.00244	−0.01262	−0.00107
	15	0.55	0.26997	−0.09700	−0.02224	−0.00062	−0.01630	0.00011
	16	0.55	0.29863	−0.09448	−0.02174	−0.00053	−0.01352	0.00198

Source: Hsiao, Pesaran, and Tahmiscioglu (1999).

finite T. However, contrary to the sampling approach where the correct modeling of initial observations is quite important, the Bayesian approach appears to perform fairly well in the estimation of the mean coefficients for dynamic random-coefficients models even when the initial observations are treated as fixed constants. The Monte Carlo study also cautions against the practice of justifying the use of certain estimators based on their asymptotic properties. Both the mean group and the corrected mean group estimators perform poorly in panels with very small T. The hierarchical Bayes estimator appears preferable to the other consistent estimators unless the time dimension of the panel is sufficiently large.

6.8 TWO EXAMPLES

6.8.1 Liquidity Constraints and Firm Investment Expenditure

The effects of financial constraints on company investment have been subject to intensive debate by economists. At one extreme, Jorgenson (1971) claims that "the evidence clearly favors the Modigliani-Miller theory (1958, 61). Internal liquidity is not an important determinant of the investment, given the level of output and external funds." At the other extreme, Stiglitz and Weiss (1981) argue that because of imperfections in the capital markets, costs of internal and external funds generally will diverge, and internal and external funds generally will not be perfect substitutes for each other. Fazzari, Hubbard, and Petersen (1988), Bond and Meghir (1994), etc. tested for the importance of internal finance by studying the effects of cash flow across different groups of companies like identifying groups of firms according to company retention practices. If the null hypothesis of perfect capital market is correct, then no difference should be found in the coefficient of cash flow variable across groups. However, these authors find that cash flow coefficient is large for companies with low dividend payout rates.

However, there is no sound theoretical basis for assuming that only low dividend payout companies are subject to financial constraints. The finding that larger companies have larger cash flow coefficients is inconsistent with both the transaction costs and asymmetric information explanations of liquidity constraints. Whether firm heterogeneity can be captured by grouping firms according to some indicators remains open to question.

Hsiao and Tahmiscioglu (1997) use COMPUSTAT annual industrial files of 561 firms in manufacturing sector for the period 1971–1992 to estimate the following five different investment expenditure models with and without using liquidity models:

$$\left(\frac{I}{K}\right)_{it} = \alpha_i^* + \gamma_i \left(\frac{I}{K}\right)_{i,t-1} + \beta_{i1}\left(\frac{LIQ}{K}\right)_{i,t-1} + \epsilon_{it}, \tag{6.8.1}$$

$$\left(\frac{I}{K}\right)_{it} = \alpha_i^* + \gamma_i \left(\frac{I}{K}\right)_{i,t-1} + \beta_{i1}\left(\frac{LIQ}{K}\right)_{i,t-1} + \beta_{i2} q_{it} + \epsilon_{it}, \tag{6.8.2}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

$$\left(\frac{I}{K}\right)_{it} = \alpha_i^* + \gamma_i \left(\frac{I}{K}\right)_{i,t-1} + \beta_{i1} \left(\frac{LIQ}{K}\right)_{i,t-1} + \beta_{i2} \left(\frac{S}{K}\right)_{i,t-1} + \epsilon_{it}, \tag{6.8.3}$$

$$\left(\frac{I}{K}\right)_{it} = \alpha_i^* + \gamma_i \left(\frac{I}{K}\right)_{i,t-1} + \beta_{i2} q_{it} + \epsilon_{it}, \tag{6.8.4}$$

and

$$\left(\frac{I}{K}\right)_{it} = \alpha_i^* + \gamma_i \left(\frac{I}{K}\right)_{i,t-1} + \beta_{i2} \left(\frac{S}{K}\right)_{i,t-1} + \epsilon_{it}. \tag{6.8.5}$$

where I_{it} is firm i's capital investment at time t, LIQ_{it} is a liquidity variable (defined as cash flow minus dividends); S_{it} is sales, q_{it} is Tobin's q (Brainard and Tobin 1968; Tobin 1969), defined as the ratio of the market value of the firm to the replacement value of capital; and K_{it} is the beginning-of-period capital stock. The coefficient β_{i1} measures the short-run impact of liquidity variable on firm i's investment in each of these three specifications. Models 4 and 5 ((6.8.4) and (6.8.5)) are two popular variants of investment equations that do not use the liquidity variable as an explanatory variable – the Tobin q model (e.g., Hayashi 1982; Summers 1981) and the sales capacity model (e.g., Kuh 1963). The sale variable can be regarded as a proxy for future demand for the firm's output. The q theory relates investment to marginal q, which is defined as the ratio of the market value of new investment goods to their replacement cost. If a firm has unexploited profit opportunities, then an increase of its capital stock of \$1 will increase its market value by more than \$1. Therefore, firm managers can be expected to increase investment until marginal q equals 1. Thus, investment will be an increasing function of marginal q. Because marginal q is unobservable, it is common in empirical work to substitute it with average or Tobin's q.

Tables 6.5 and 6.6 present some summary information from the firm by firm regressions of these five models. Table 6.5 shows the percentage of significant coefficients at the 5 percent significance level for a one-tailed test. Table 6.6 shows the first and third quartiles of the estimated coefficients. The estimated coefficients vary widely from firm to firm. The F-test of slope homogeneity across firms while allowing for firm-specific intercepts is also rejected (see Table 6.5).

The approach of relating the variation of β_{i1} to firm characteristics such as dividend payout rate, company size, sales growth, capital intensity, standard deviation of retained earnings, debt-to-equity ratio, measures of liquidity stocks from the balance sheet, number of shareholders, and industry dummies is unsuccessful. These variables as a whole do not explain the variation of estimated β_{i1} well. The maximum $\bar{R}^2$ is only 0.113. Many of the estimated coefficients are not significant under various specifications. Neither can one substitute functions of the form (6.5.2) into (6.8.1)–(6.8.5) and estimate the coefficients directly because of perfect multicollinearity. So Hsiao and Tahmiscioglu (1997) classify firms into reasonably homogeneous groups using the

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Table 6.5. *Individual firm regressions (percentage of firms with significant coefficients)*

	Percentage of firms				
	Model 1	2	3	4	5
Coefficient for:					
$(\mathrm{LIQ}/K)_{t-1}$	46	36	31		
q		31		38	
$(S/K)_{t-1}$			27		44
Percentage of firms with significant autocorrelation	14	12	13	20	15
Actual F	2.47	2.98	2.01	2.66	2.11
Critical F	1.08	1.08	1.08	1.06	1.06

Note: The number of firms is 561. The significance level is 5 percent for a one-tailed test. Actual F is the F statistic for testing the equality of slope coefficients across firms. For the F test, the 5 percent significance level is chosen. To detect serial correlation, Durbin's t-test at the 5 percent significance level is used.

Source: Hsiao and Tahmiscioglu (1997, Table 1).

capital intensity ratio of 0.55 as a cut-off point. Capital intensity is defined as the minimum value of the ratio of capital stock to sales over the sample period. It is the most statistically significant and most stable variable under different specifications.

Table 6.7 presents the variable intercept estimates for the groups of less and more capital intensive firms. The liquidity variable is highly significant in all three variants of the liquidity model. There are also significant differences in the coefficients of the liquidity variable across the two groups. However, Table 6.7 also shows that the null hypothesis of the equality of slope coefficients conditioning on the firm-specific effects is strongly rejected for all specifications

Table 6.6. *Coefficient heterogeneity: slope estimates at first and third quartiles across a sample of 561 firms*

	Slope estimates			
Model	$(I/K)_{i,t-1}$	$(\mathrm{LIQ}/K)_{i,t-1}$	q_{it}	$(S/K)_{i,t-1}$
1	.026, .405	.127, .529		
2	−.028, .359	.062, .464	0, .039	
3	.100, .295	.020, .488		−.005, .057
4	.110, .459		.007, .048	
5	−.935, .367			.012, .077

Source: Hsiao and Tahmiscioglu (1997, Table 2).

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Table 6.7. *Variable intercept estimation of models for less- and more-capital-intensive firms*

Variable	Variable intercept estimate					
	Less-capital-intensive firms			More-capital-intensive firms		
$(I/K)_{i,t-1}$	.265	.198	.248	.392	.363	.364
	(.011)	(.012)	(.011)	(.022)	(.023)	(.022)
$(\text{LIQ}/K)_{i,t-1}$	.161	.110	.119	.308	.253	.278
	(.007)	(.007)	(.007)	(.024)	(.027)	(.025)
$(S/K)_{i,t-1}$		.023			.025	
		(.001)			(.006)	
q_{it}			.011			.009
			(.0006)			(.002)
Actual F	2.04	1.84	2.22	2.50	2.19	2.10
Critical F	1.09	1.07	1.07	1.20	1.17	1.17
Numerator d.f.	834	1,251	1,251	170	255	255
Denominator d.f.	6,592	6,174	6,174	1,368	1,282	1,282
Number of firms	418	418	418	86	86	86

Note: The dependent variable is $(I/K)_{it}$. Less-capital-intensive firms are those with minimum (K/S) between 0.15 and 0.55 over the sample period. For more-capital-intensive firms, the minimum (K/S) is greater than 0.55. The regressions include company-specific intercepts. Actual F is the F statistic for testing the homogeneity of slope coefficients. For the F test, a 5 percent significance level is chosen. The estimation period is 1974–1992. Standard errors are in parentheses.
Source: Hsiao and Tahmiscioglu (1997, Table 5).

for both groups. In other words, using the capital intensity ratio of 0.55 as a cut-off point, there is still substantial heterogeneity within the groups.

As neither there appears to have a set of explanatory variables that adequately explains the variation of β_{i1}, nor can homogeneity be achieved by classifying firms into groups, one is left with either treating $\boldsymbol{\beta}_i$ as fixed and different or treating $\boldsymbol{\beta}_i$ as random draws from a common distribution. Within the random-effects framework, individual differences are viewed as random draws from a population with constant mean and variance. Therefore, it is appropriate to pool the data and try to draw some generalization about the population. On the other hand, if individual differences reflect fundamental heterogeneity or if individual response coefficients depend on the values of the included explanatory variables, estimation of the model parameters based on the conventional random effects formulation can be misleading. To avoid this bias, heterogeneity among individuals must be treated as fixed. In other words, one must investigate investment behavior firm by firm, and there is no advantage of pooling. Without pooling, the shortage of degrees of freedom and multicollinearity can render the resulting estimates meaningless and make drawing general conclusions difficult.

Table 6.8 presents the estimates of the mixed fixed- and random-coefficients model of the form (6.6.24) by assuming that conditional on company-specific effects, the remaining slope coefficients are randomly distributed around a

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Table 6.8. *Estimation of mixed fixed- and random-coefficient models for less- and more-capital-intensive firms*

	Estimate					
Variable	Less-capital-intensive firms			More-capital-intensive firms		
$(I/K)_{i,t-1}$	.230	.183	.121	.321	.302	.236
	(.018)	(.017)	(.019)	(.036)	(.037)	(.041)
$(\text{LIQ}/K)_{i,t-1}$	.306	.252	.239	.488	.449	.416
	(.021)	(.023)	(.027)	(.065)	(.067)	(.079)
$(S/K)_{i,t-1}$		.024			.038	
		(.003)			(.015)	
q_{it}		.019			.022	
		(.003)			(.008)	
Number of firms	418	418	418	86	86	86

Note: The dependent variable is $(I/K)_{it}$. The regressions include fixed firm-specific effects. The estimation period is 1974–1992. Standard errors are in parentheses.
Source: Hsiao and Tahmiscioglu (1997, Table 7).

certain mean within each of less and more capital-intensive groups. To evaluate the appropriateness of these specifications, Table 6.9 presents the comparison of the recursive predictive density of the mixed fixed- and random-coefficients models and the fixed-coefficients model assuming that each company has different coefficients for the three variants of the liquidity model by dividing

Table 6.9. *Least-squares estimation of aggreate money demand function*

Dependent variable	Sample period	Variable	Parameter estimate	Standard error
M2	1980.IV–2000.IV	Intercept	1.30462	0.28975
		Real GDP	−0.15425	0.04538
		RM2(-1)	1.07022	0.02790
		Bond rate	−0.00186	0.00069
	1992.I–2000.IV	Intercept	−0.16272	0.85081
		Real GDP	0.00847	0.06772
		RM2(-1)	1.00295	0.02248
		Bond rate	−0.00250	0.00140
M1	1980.IV–2000.IV	Intercept	0.46907	0.21852
		Real GDP	−0.01857	0.01700
		RM(-1)	0.98964	0.01249
		Bond rate	−0.00566	0.00135
	1992.I–2000.IV	Intercept	−0.68783	2.10228
		Real GDP	0.08414	0.14898
		RM1(-1)	0.96038	0.019999
		Bond rate	−0.01005	0.00283

Source: Hsiao, Shen and Fujiki (2005, Table 5).

the sample into pre- and post-1989 periods. The numbers reported in Table 6.8 are the logarithms of (6.6.28). The results indicate that the mixed fixed- and random-coefficients model is favored over the fixed-coefficients model for both groups. Similar comparisons between the liquidity model, Tobin's q, and sales accelerator models also favor liquidity as an important explanatory variable.

Table 6.8 shows that the estimated liquidity coefficients are highly significant and there are significant differences between different classes of companies. The mean coefficient of the liquidity variable turns out to be 60 to 80 percent larger for the more capital-intensive group than for the less capital-intensive group. The implied long-run relationships between the liquidity variable and the fixed investment variable are also statistically significant. For instance, for model (6.8.1), a 10% increase in liquidity capital ratio leads to a 4% increase in fixed investment capital ratio in the long run for the less capital-intensive group compared to a 7% increase in the ratio for the more capital-intensive group. The mixed model also yields substantially larger coefficient estimates of the liquidity variable than those obtained from the variable intercept model. If the coefficients are indeed randomly distributed and the explanatory variables are positively autocorrelated, then this is precisely what one would expect from the within-estimates (Pesaran and Smith 1995).

In short, there are substantial differences across firms in their investment behavior. When these differences are ignored by constraining the parameters to be identical across firms, the impact of liquidity variable on firm investment is seriously underestimated. The mixed fixed- and random-coefficients model appears to fit the data well. The mixed model allows pooling and allows some general conclusions to be drawn about a group of firms. The estimation results and prediction tests appear to show that financial constraints are the most important factor affecting actual investment expenditure, at least for a subset U.S. manufacturing companies.

6.8.2 Aggregate versus Disaggregate Analysis

A model is a simplification of the real world. The purpose is to capture the essential factors that affect the outcomes. One of the tools for reducing the real-world detail is through "suitable" aggregation. However, for aggregation not to distort the fundamental behavioral relations among economic agents, certain "homogeneity" conditions must hold between the micro-units. Many economists have shown that if micro-units are heterogeneous, aggregation can lead to very different relations among macro-variables from those of the micro-relations (e.g., Lewbel 1992, 1994; Pesaran 2003; Stoker 1993; Theil 1954; Trivedi 1985).

For instance, consider the simple dynamic equation,

$$y_{it} = \gamma_i y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta}_i + \alpha_i + u_{it}, \ | \gamma_i | < 1, \quad i = 1, \ldots, N, \quad (6.8.6)$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

where the error u_{it} is covariance stationary. Equation (6.8.6) implies a long-run relation between y_{it} and $\mathbf{x}_{it}$,

$$y_{it} - \mathbf{x}_{it}'\mathbf{b}_i - \eta_i = v_{it} \tag{6.8.7}$$

where $\mathbf{b}_i = (1-\gamma_i)^{-1}\boldsymbol{\beta}_i$, $\eta_i = (1-\gamma_i)^{-1}\alpha_i$, $v_{it} = (1-\gamma_i)^{-1}u_{it}$.

Let $y_t = \sum_{i=1}^N y_{it}$ and $\mathbf{x}_t = \sum_{i=1}^N \mathbf{x}_{it}$; then a similar long-run relation between $\mathbf{y}_t$ and $\mathbf{x}_t$,

$$y_t - \mathbf{x}_t'\mathbf{b} - c = v_t, \tag{6.8.8}$$

holds for a stationary v_t if and only if either of the following conditions hold (Hsiao, Shen, and Fujiki 2005):

(1) $\frac{1}{1-\gamma_i}\boldsymbol{\beta}_i = \frac{1}{1-\gamma_j}\boldsymbol{\beta}_j$ for all i and j; or

(2) if $\frac{1}{1-\gamma_i}\boldsymbol{\beta}_i \neq \frac{1}{1-\gamma_j}\boldsymbol{\beta}_j$, then $\mathbf{x}_t' = (\mathbf{x}_{1t}', \ldots, \mathbf{x}_{Nt}')$ must lie on the null space of D for all t, where $D' = \left(\frac{1}{1-\gamma_1}\boldsymbol{\beta}_1' - \mathbf{b}', \ldots, \frac{1}{1-\gamma_N}\boldsymbol{\beta}_N' - \mathbf{b}'\right)$.

Panel data provide information on micro-units. They can be used to check if either of these two suitable aggregation conditions hold. For instance, Hsiao, Shen, and Fujiki (2005) find that the estimated aggregate relations between (real) money demand, (real) GDP, and (5-year) bond rate are unstable and sensitive to the time period dovered (see Table 6.9). Depending on the sample period covered, the estimated relations are either of wrong sign or statistically insignificant. They find that the estimated long-run income elasticities are 75.23 for M1 and 11.04 for M2, respectively, an incredible magnitude.

Hsiao, Shen, and Fujiki (2005) attribute the "incredible" results from aggregate data analysis to the "heterogeneity" among the 47 prefectures of Japan. When micro-relations are "heterogeneous," one way is to estimate each micro-relation separately. However, there may not have enough time series observations to obtain reliable micro-relations. Moreover, policymakers are interested in average relations, not individual relations. A random coefficienet framework is a convenient formulation that take account indivdiual heterogeneity while still allowing the estimation of average relation. Table 6.10 provides a random coefficient model estimates of the mean relation between (real) money demand and (real) GDP and (5-year) bond rate for teh 40 Japanese perfectures. The estimated short-run income elasticity for M1 and M2 is 0.88 and 0.47, respectively. The long-run income elasticity is 2.56 for M1 and 1.01 for M2. These results appear to be consistent with economic theory and the broadly observed facts about Japan. The average growth rate for M2 in the 1980s is about 9.34 percent. The inflation rate is 1.98 percent. The real M2 growth rate is 7.36 percent. The real growth rate of GDP during this period is 4.13 percent. Taking account the impact of 5-year bond rate fell from 9.332 percent at 1980.I to 5.767 at 1989.IV, the results are indeed very close to the estimated long-run income elasticites based on disaggregate data analysis.

If "heterogeneity" is indeed present in micro-units, then shall we predict the aggregate outcome based on the summation of estimated micro-relations

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

Table 6.10. *Random-coefficient estimates of Japan Prefecture money demand equation*

	M1		M2	
	Coefficient	Standard error	Coefficient	Standard error
Lagged money	0.656	0.034	0.533	0.069
Income	0.881	0.114	0.473	0.064
Bond Rate	−0.0476	0.006	−0.009	0.003
Constant	−2.125	0.038	0.043	0.239
	Variance–covariance matrix of $M1(\gamma_i, \boldsymbol{\beta}_i')$			
	0.015			
	−0.001	0.177		
	0.001	−0.059	0.0005	
	−0.024	−0.588	−0.023	2.017
	Variance–covariance matrix of $M2$ equation $(\gamma_i, \boldsymbol{\beta}_i')$.			
	0.068			
	−0.031	0.062		
	0.002	0.0003	0.0014	
	−0.13	−0.107	−0.009	0.8385

Source: Hsiao, Shen and Fujiki (2005, Table 1).

or shall we predict the aggregate outcomes based on the estimated aggregate relations? Unfortunately, there is not much work on this specific issue. In choosing between whether to predict aggregate variables using aggregate (H_a) or disaggregate (H_d) equations, Griliches and Grunfeld (1960) suggest using the criterion of:

$$\text{Choose } H_d \text{ if } \mathbf{e}_d'\mathbf{e}_d < \mathbf{e}_a'\mathbf{e}_a; \quad \text{otherwise choose } H_a$$

where $\mathbf{e}_d$ and $\mathbf{e}_a$ are the estimates of the errors in predicting aggregate outcomes under H_d and H_a, respectively. Hsiao, Shen, and Fujiki (2005) provide a simulation comparison on artificially generated time series data for each prefecture based on the observed stylized facts. Table 6.11 presents the

Table 6.11. *Error sum of squares (ESS) and predicted error sum of squares (PES) for disaggregate and aggregate data*

	M1		M2	
	Aggregate data	Disaggregate data	Aggregate data	Disaggregate data
EES	3.78×10^9	1.35×10^6	3.59×10^{43}	7.45×10^{42}
PES	2.51×10^{10}	5.75×10^7	9.55×10^{45}	2.04×10^{43}

Source: Hsiao, Shen and Fujiki (2005, Table VIII).

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

within-sample fit comparisons in the first row and the post-sample prediction comparison in the second row. Both criteria unambiguously favour predicting aggregate outcomes by summing the outcomes from the disaggregate equations.

6.9 CORRELATED RANDOM-COEFFICIENTS MODELS

6.9.1 Introduction

Standard random-coefficients models assume the variation of coefficients are independent of the variation of regressors (e.g., Chapter 6, Section 6.2.1; Hsiao and Pesaran 2009). In recent years, a great deal of attention has been devoted to the correlated random coefficients model (e.g., Card 1995; Heckman and Vytlacil 1998; Heckman, Urzua, and Vytlacil, 2006; Heckman, Schmierer, and Urzua 2010). This type of model is motivated by the measurement of treatment effect of a policy. For instance, in the study of return to schooling, it is plausible that there are unmeasured ability or motivation factors that affect the return to schooling and are also correlated with the level of schooling (e.g., Card 1995; Heckman and Vytlacil 1998). As a matter of fact, Li and Tobias (2011) find strong evidence that the amount of schooling attained is determined, in part, by the individual's own return to education. Specifically a one percentage increase in the return to schooling is associated with roughly 0.2 more years of education.

A common formulation for a correlated random-coefficients model is to let

$$\boldsymbol{\beta}_i = \bar{\boldsymbol{\beta}} + \boldsymbol{\alpha}_i. \tag{6.9.1}$$

Substituting (6.9.1) into the regression model (6.1.2) yields

$$y_{it} = \mathbf{x}_{it}'\bar{\boldsymbol{\beta}} + \mathbf{x}_{it}'\boldsymbol{\alpha}_i + u_{it}, \tag{6.9.2}$$

where

$$E\boldsymbol{\alpha}_i = \mathbf{0}, \tag{6.9.3}$$

$$E\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j' = \begin{cases} \Delta, & \text{if } i = j. \\ \mathbf{0}, & \text{if } i \neq j. \end{cases} \tag{6.9.4}$$

and

$$E(u_{it} \mid \mathbf{x}_{it}, \boldsymbol{\beta}_i) = 0, \tag{6.9.5}$$

However, we now assume

$$E\mathbf{x}_{it}\boldsymbol{\alpha}_i' \neq \mathbf{0}. \tag{6.9.6}$$

Let $v_{it} = \mathbf{x}_{it}'\alpha_i + u_{it}$; then

$$E(v_{it} \mid \mathbf{x}_{it}) \neq 0. \tag{6.9.7}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

6.9.2 Identification with Cross-Sectional Data

If only cross-sectional observations of $(y, \mathbf{x})$ are available, it is not possible to identify $\bar{\boldsymbol{\beta}}$. Nor does the existence of instruments $\mathbf{z}_1$ such that

$$\operatorname{cov}(\mathbf{z}_1, \mathbf{x}) \neq \mathbf{0}, \tag{6.9.8}$$

$$\operatorname{cov}(\mathbf{z}_1, u) = \mathbf{0} \tag{6.9.9}$$

alone is sufficient to identify $\bar{\boldsymbol{\beta}}$ because the instrumental variable estimator

$$\begin{aligned}
\hat{\bar{\boldsymbol{\beta}}}_{iv} &= \left[\left(\sum_{i=1}^{N} \mathbf{x}_i \mathbf{z}_{1i}'\right)\left(\sum_{i=1}^{N} \mathbf{z}_{1i}\mathbf{z}_{1i}'\right)^{-1}\left(\sum_{i=1}^{N} \mathbf{z}_{1i}\mathbf{x}_i\right)\right]^{-1} \\
&\quad \cdot \left[\left(\sum_{i=1}^{N} \mathbf{x}_i \mathbf{z}_{1i}'\right)\left(\sum_{i=1}^{N} \mathbf{z}_{1i}\mathbf{z}_{1i}'\right)^{-1}\left(\sum_{i=1}^{N} \mathbf{z}_{1i} y_i\right)\right] \\
&= \bar{\boldsymbol{\beta}} + \left[\left(\sum_{i=1}^{N} \mathbf{x}_i \mathbf{z}_{1i}'\right)\left(\sum_{i=1}^{N} \mathbf{z}_{1i}\mathbf{z}_{1i}'\right)^{-1}\left(\sum_{i=1}^{N} \mathbf{z}_{1i}\mathbf{x}_i'\right)\right]^{-1} \\
&\quad \cdot \left[\left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i \mathbf{z}_{1i}'\right)\left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_{1i}\mathbf{z}_{1i}'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_{1i}\mathbf{x}_i'\boldsymbol{\alpha}_i + \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_{1i}u_i\right)\right].
\end{aligned} \tag{6.9.10}$$

Although under (6.9.9) $\operatorname{plim} \frac{1}{N}\sum_{i=1}^{n} \mathbf{z}_{1i}u_i = \mathbf{0}$,

$$\begin{aligned}
&\operatorname{plim} \frac{1}{N}\sum_{i=1}^{n} \mathbf{z}_{1i}\mathbf{x}_i'\boldsymbol{\alpha}_i \\
&\quad = E[\mathbf{z}_1 E(\mathbf{x}'\boldsymbol{\alpha} \mid \mathbf{z}_1)] \\
&\quad = E[\mathbf{z}_1 E(\mathbf{x}' \mid \mathbf{z}_1) E(\boldsymbol{\alpha} \mid \mathbf{x}, \mathbf{z}_1)],
\end{aligned} \tag{6.9.11}$$

which is not equal 0 given (6.9.8) and the assumption that $E(\boldsymbol{\alpha} \mid \mathbf{x}) \neq \mathbf{0}$.

To identify $\bar{\boldsymbol{\beta}}$, the variation of $\mathbf{x}_i$ and $\boldsymbol{\beta}_i$ need to be independent conditional on $\mathbf{z}_{1i}$. In other words, we need exclusion restrictions. Heckman and Vytlacil (1998) consider estimating $\bar{\boldsymbol{\beta}}$ assuming the existence of instruments $\mathbf{z}_i = (\mathbf{z}_{1i}, \mathbf{z}_{2i})$ such that

$$\mathbf{x}_i = \Pi \mathbf{z}_{1i} + \mathbf{v}_i \tag{6.9.12}$$

$$\boldsymbol{\beta}_i = \Phi \mathbf{z}_{2i} + \boldsymbol{\eta}_i \tag{6.9.13}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

where $\mathbf{z}_{1i}$ and $\mathbf{z}_{2i}$ are $m_1 \times 1$ and $m_2 \times 1$ vectors of instruments that satisfy

$$E(u_i \mid \mathbf{z}_i) = 0, \tag{6.9.14}$$

$$E(\boldsymbol{\eta}_i \mid \mathbf{z}_i) = \mathbf{0}, \tag{6.9.15}$$

$$E(\mathbf{v}_i \mid \mathbf{z}_i) = \mathbf{0}, \tag{6.9.16}$$

and $\mathbf{z}_2$ contains elements that are not in $\mathbf{z}_1$.

Then, under (6.9.13), if Φ is known, an estimator for $\bar{\boldsymbol{\beta}}$ can be obtained from the relation,

$$E(\boldsymbol{\beta}) = \bar{\boldsymbol{\beta}} = \Phi E(\mathbf{z}_2). \tag{6.9.17}$$

Substituting (6.9.12) and (6.9.13) into (6.9.14) yields

$$\begin{aligned} y_i &= \mathbf{x}_i'\boldsymbol{\beta}_i + u_i \\ &= (\mathbf{z}_{1i}'\Pi' + \mathbf{v}_i')(\Phi\mathbf{z}_{2i} + \boldsymbol{\eta}_i) + u_i \\ &= (\mathbf{z}_{1i}'\boldsymbol{\pi}_1\mathbf{z}_{2i}')\boldsymbol{\phi}_1 \ + \ (\mathbf{z}_{1i}'\boldsymbol{\pi}_2\mathbf{z}_{2i}')\boldsymbol{\phi}_2 \\ &\quad + \cdots + (\mathbf{z}_{1i}'\boldsymbol{\pi}_K\mathbf{z}_{2i}')\boldsymbol{\phi}_K \ + \ E(\mathbf{v}_i'\boldsymbol{\eta}_i \mid \mathbf{z}_i) + \epsilon_i^*, \end{aligned} \tag{6.9.18}$$

where $\boldsymbol{\pi}_k'$ and $\boldsymbol{\phi}_k'$ denote the kth row of Π and Φ, respectively,

$$\boldsymbol{\epsilon}_i^* = \mathbf{v}_i'\Phi\mathbf{z}_{2i} + \boldsymbol{\eta}_i'\Pi\mathbf{z}_{1i} + [\mathbf{v}_i'\boldsymbol{\eta}_i - E(\mathbf{v}_i'\boldsymbol{\eta}_i \mid \mathbf{z}_i)] + u_i. \tag{6.9.19}$$

Under (6.9.12)–(6.9.16), $E(\boldsymbol{\epsilon}_i^* \mid \mathbf{z}_i) = 0$.

Therefore, a consistent estimator of Φ exists if

$$E(\mathbf{v}_i\boldsymbol{\eta}_i' \mid \mathbf{z}_{1i}, \mathbf{z}_{2i}) = \Sigma_{v\eta} \tag{6.9.20}$$

is not a function of $\mathbf{z}_{1i}$ and $\mathbf{z}_{2i}$, and

$$\operatorname{rank}\left[\frac{1}{N}\sum_{i=1}^{n}\left(\mathbf{z}_{2i}\mathbf{z}_{1i}'\otimes\hat{\Pi}\right)\left(\hat{\Pi}'\otimes\mathbf{z}_{1i}\mathbf{z}_{2i}'\right)\right] = Km_2. \tag{6.9.21}$$

In other words, the necessary condition for the identification of $E(\boldsymbol{\beta}_i) = \bar{\boldsymbol{\beta}}$ for the correlated random-coefficients model (6.9.1)–(6.9.7) when only cross-sectional data are available are that there exist m_1 instruments for $\mathbf{x}_i$ and nonzero m_2 instruments for $\boldsymbol{\beta}_i$ that satisfy (6.9.12)–(6.9.16) with $m_1^2 > Km_2, m_2 > 0$, and either (6.9.20) holds or $E(\mathbf{v}_i'\boldsymbol{\eta}_i \mid \mathbf{z}_i)$ is known.

The requirements that there exist nonzero $\mathbf{z}_1$ and $\mathbf{z}_2$ with $m_1^2 \geq Km_2$, and (6.9.20) holds are stronger than the usual requirement for the existence of an instrumental variable estimator. As a matter of fact, the necessary condition requires the existence of both $\mathbf{z}_1$ and $\mathbf{z}_2$ (i.e., $m_1 > 0, m_2 > 0$). Neither is (6.9.20) an innocuous assumption. To the best of my knowledge, the conditional covariance independent of $\mathbf{z}$ holds only if $\mathbf{v}$, $\boldsymbol{\eta}$, and $\mathbf{z}$ are joint normal.

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

6.9.3 Estimation of the Mean Effects with Panel Data

When only cross-sectional data are available, the identification conditions of average effects for a correlated random-coefficients model are very stringent and may not be satisfied for many data sets. The instrumental variable approach requires the estimation of a large number of parameters $[(m_1 + m_2)K]$. Multicollinearity and shortages of degrees of freedom could lead to very unreliable estimates. On the other hand, panel data, by blending interindividual differences with intraindividual dynamics can offer several alternatives to get around the difficulties of the correlations between the coefficients and the regressors without the prior conjecture of the existence of certain instruments that satisfy the exclusion restrictions. For ease of exposition, we suppose there are T time series observations of $(y_{it}, \mathbf{x}_{it})$ for each individual i. Let $(\mathbf{y}_i', \mathbf{x}_i')$ be the stacked T time series observations of y_{it} and $\mathbf{x}_{it}'$ for each i.

6.9.3.1 Group Mean Estimator

We note that condition on $\mathbf{x}_i, \boldsymbol{\beta}_i$ is a fixed constant. Under (6.9.5), the least-squares estimator of the equation

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}_i + u_{it}, \quad t = 1, \ldots, T, \tag{6.9.22}$$

yields an unbiased estimator of $\boldsymbol{\beta}_i$, $\hat{\boldsymbol{\beta}}_i$, for each i with covariance matrix $\sigma_i^2(X_i'X_i)^{-1}$ if u_{it} is independently distributed over t, where X_i denotes the $T \times K$ stacked $(\mathbf{x}_{it}')$. If u_{it} is independently distributed over i, taking the simple average of $\hat{\boldsymbol{\beta}}_i$ as in Hsiao, Pesaran, and Tahmiscioglu (1999),

$$\hat{\bar{\boldsymbol{\beta}}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i. \tag{6.9.23}$$

yields a consistent estimator of $\bar{\boldsymbol{\beta}}$ as $N \to \infty$. If $T > K$, the estimator (6.9.23) is consistent and asymptotically normally distributed as $N \to \infty$, and $\sqrt{N}(\hat{\bar{\boldsymbol{\beta}}} - \bar{\boldsymbol{\beta}})$ is asymptotically normally distributed with mean $\mathbf{0}$ and covariance matrix

$$\text{Asy Cov}\,(\hat{\bar{\boldsymbol{\beta}}}) = \left[\Delta + \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2(X_i'X_i)^{-1}\right], \tag{6.9.24}$$

if u_{it} is independently distributed over i and t with variance σ_i^2.

6.9.3.2 Conventional Fixed-Effects Estimator

The estimator (6.9.23) is simple to implement. However, if $T < K$, we cannot estimate $\boldsymbol{\beta}_i$ using the ith individual's time series observations $(y_i, \mathbf{x}_i')$. Nevertheless, the conventional fixed-effects estimator can still allow us to obtain consistent estimator of $\bar{\boldsymbol{\beta}}$ in a number of situations.

Let $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $\bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it}$. The conventional fixed-effects estimator first takes the deviation of each observation from its time series mean,

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

and then regresses $(y_{it} - \bar{y}_i)$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$ (e.g., Chapter 3). Model (6.9.2) leads to

$$(y_{it} - \bar{y}_i) = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\bar{\boldsymbol{\beta}} + (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\alpha}_i + (u_{it} - \bar{u}_i), \begin{matrix} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{matrix} \quad (6.9.25)$$

where $\bar{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}$.

The fixed-effects estimator (6.9.25) will converge to $\bar{\boldsymbol{\beta}}$ provided

$$\text{plim } \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\alpha}_i = \mathbf{0} \quad (6.9.26)$$

and

$$\text{plim } \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(u_{it} - \bar{u}_i) = \mathbf{0} \quad (6.9.27)$$

Under (6.9.5), (6.9.27) holds. Under the assumption that $\mathbf{x}_{it}$ and $\boldsymbol{\alpha}_i$ are linearly related with finite variance, then $\mathbf{x}_{it} - \bar{\mathbf{x}}_i$ does not involve $\boldsymbol{\alpha}_i$, and hence (6.9.26) holds by a law of large numbers. Hence the conventional fixed-effects estimator is $\sqrt{N}$ consistent and asymptotically normally distributed as $N \to \infty$. The asymptotic covariance matrix of the conventional fixed-effects estimator (3.2.5) can be approximated using the Newey–West heteroscedasticity-autocorrelation consistent formula (Vogelsang 2012).

When $(\mathbf{x}_{it}, \boldsymbol{\alpha}_i)$ jointly have an elliptical distribution[21] (e.g., Fang and Zhang 1990; Gupta and Varga 1993), $\mathbf{x}_{it}$ and $\boldsymbol{\alpha}_i$ are linearly related. Another case that the fixed-effects estimator can be consistent is that $(\mathbf{x}_{it}, \boldsymbol{\alpha}_i)$ are jointly symmetrically distributed; then $\frac{1}{NT}\sum_{i=1}^{N}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\alpha}_i$ will converge to 0 even though $\mathbf{x}_{it}$ have a mean different from 0.

6.9.3.3 Panel Pooled Least-Squares Estimator

The conventional fixed-effects (FE) estimator (3.2.5) can yield a consistent estimator of $\bar{\boldsymbol{\beta}}$, under certain conditions. However, if $\boldsymbol{\alpha}$ and $\mathbf{x}_{it}$ are not linearly related, it is inconsistent. Moreover, if $\mathbf{x}_{it}$ contains time-invariant variables, then the mean effects of time-invariant variables cannot be identified by the conventional fixed-effects estimator. Furthermore, the FE estimator only makes use of within- (group) variation. Because in general the between-group variation is much larger than within-group variation, the FE estimator could also mean a loss of efficiency. To get around these limitations on the FE estimator as well as allowing the case that $\boldsymbol{\alpha}_i$ and $\mathbf{x}_{it}$ are not linearly related, Hsiao, Li, Liang, and Xie (2012) suggest a modified specification to obtain the estimate of the mean effects.

[21] Many commonly assumed distributions such as uniform, normal, Student's t, double exponential, etc. belong to the family of elliptical distributions.

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

To illustrate the basic idea, we first assume that $E(\alpha_i \mid \mathbf{x}_i)$ is a linear function of $\mathbf{x}_i$. We will show later that a similar procedure can be applied if $E(\boldsymbol{\alpha}_i \mid \mathbf{x}_i)$ is a function of a higher order of $\mathbf{x}_i$.

A6.9.1: $(\mathbf{x}_i', \boldsymbol{\alpha}_i')$ are independently, identically distributed across i with

$$E(\boldsymbol{\alpha}_i \mid \mathbf{x}_i) = \mathbf{a} + B\mathbf{x}_i, \tag{6.9.28}$$

where $\mathbf{a}$ and B are the $K \times 1$ and $K \times TK$ constant vector and matrix, respectively.

From (6.9.3) and (6.9.28), we have

$$E_x[E(\boldsymbol{\alpha}_i \mid \mathbf{x}_i)] = \mathbf{a} + BE(\mathbf{x}_i) = \mathbf{0}. \tag{6.9.29}$$

It follows that

$$E(\boldsymbol{\alpha}_i \mid \mathbf{x}_i) = B(\mathbf{x}_i - E\mathbf{x}_i) \tag{6.9.30}$$

Substituting

$$\boldsymbol{\alpha}_i = E(\boldsymbol{\alpha}_i \mid \mathbf{x}_i) + \boldsymbol{\omega}_i \tag{6.9.31}$$

and (6.9.30) into (6.9.1) yields

$$y_{it} = \mathbf{x}_{it}'\bar{\boldsymbol{\beta}} + \mathbf{x}_{it}'B(\mathbf{x}_i - E\mathbf{x}_i) + v_{it}^*, \tag{6.9.32}$$

where

$$v_{it}^* = \mathbf{x}_{it}'\boldsymbol{\omega}_i + u_{it}. \tag{6.9.33}$$

By construction, $E(v_{it}^* \mid \mathbf{x}_i) = 0$. Therefore, the least-squares regression of

$$\mathbf{y}_i = X_i\bar{\boldsymbol{\beta}} + X_i \otimes (\mathbf{x}_i - \bar{\mathbf{x}})' \operatorname{vec}(B') + \mathbf{v}_i \tag{6.9.34}$$

yields $\sqrt{N}$ consistent and asymptotically normally distributed estimator of $\bar{\boldsymbol{\beta}}$ when $N \to \infty$, where $\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i$, and $\mathbf{v}_i^* = (v_{i1}^*, \ldots, v_{iT}^*)'$ as long as

$$\frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix} X_i'X_i & X_i'X_i \otimes (\mathbf{x}_i' - \bar{\mathbf{x}}') \\ (\mathbf{x}_i - \bar{\mathbf{x}}) \otimes X_i'X_i & (\mathbf{x}_i - \bar{\mathbf{x}}) \otimes X_i'X_i \otimes (\mathbf{x}_i - \bar{\mathbf{x}})' \end{bmatrix} \tag{6.9.35}$$

is a full rank matrix.

However, because

$$E\mathbf{v}_i\mathbf{v}_i' = X_i\Delta^*X_i' + \sigma_i^2 I_T, \tag{6.9.36}$$

where $\Delta^* = E(\boldsymbol{\omega}_i\boldsymbol{\omega}_i')$, a more efficient estimator of $\bar{\boldsymbol{\beta}}$ will be a generalized least-squares estimator (GLS) if Δ^* and σ_i^2 are known. If Δ^* and σ_i^2 are unknown, we can apply the feasible GLS (FGLS) through a two-step procedure.

Similar reasoning can be applied if $E(\boldsymbol{\alpha}_i \mid \mathbf{x}_i)$ is a higher order polynomial of $\mathbf{x}_i$, say

$$E(\boldsymbol{\alpha}_i \mid \mathbf{x}_i) = \mathbf{a} + B\mathbf{x}_i + C\mathbf{x}_i \otimes \mathbf{x}_i. \tag{6.9.37}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

then from $E_x[E(\boldsymbol{\alpha}_i \mid \mathbf{x}_i)] = 0$, it follows that then the least-squares regression of

$$\begin{aligned} y_{it} = \mathbf{x}_{it}'\bar{\boldsymbol{\beta}} &+ \mathbf{x}_{it}' \otimes (\mathbf{x}_i - E\mathbf{x}_i)' \operatorname{vec}(B') \\ &+ \mathbf{x}_{it}' \otimes [(\mathbf{x}_i \otimes \mathbf{x}_i) - E(\mathbf{x}_i \otimes \mathbf{x}_i)]' \operatorname{vec}(C') \\ &+ v_{it}, \end{aligned} \tag{6.9.38}$$

is consistent when $N \to \infty$, where $v_{it} = \mathbf{x}_{it}'\boldsymbol{\omega}_i + u_{it}$ and $\boldsymbol{\omega}_i = \boldsymbol{\alpha}_i - E(\boldsymbol{\alpha}_i \mid \mathbf{x}_i)$.

LS (or FGLS) regression of (6.9.38) not only requires the estimation of a large number of parameters, but it could also raise the issue of multicollinearity. However, if $\mathbf{x}_{it}$ is stationary, a more parsimonious approximation would be to follow Mundlak (1978a) to replace (6.9.37) by

$$E(\boldsymbol{\alpha}_i \mid \mathbf{x}_i) = \mathbf{a} + \tilde{B}\bar{\mathbf{x}}_i + \tilde{C}(\bar{\mathbf{x}}_i \otimes \bar{\mathbf{x}}_i), \tag{6.9.39}$$

where $\bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it}$ and regressing

$$\begin{aligned} y_{it} = \mathbf{x}_{it}'\bar{\boldsymbol{\beta}} &+ \mathbf{x}_{it}' \otimes (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \operatorname{vec}(\tilde{B}') \\ &+ \mathbf{x}_{it}' \otimes [(\bar{\mathbf{x}}_i \otimes \bar{\mathbf{x}}_i) - (\bar{\mathbf{x}} \otimes \bar{\mathbf{x}})]' \operatorname{vec}(\tilde{C}') \\ &+ v_{it}, \end{aligned} \tag{6.9.40}$$

where $\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i$, $(\bar{\mathbf{x}} \otimes \bar{\mathbf{x}}) = \frac{1}{N}\sum_{i=1}^{N}[\bar{\mathbf{x}}_i \otimes \bar{\mathbf{x}}_i]$.

Remark 6.9.1: When $\mathbf{x}_{it}$ contains an intercept term, (6.9.35) is not a full rank matrix. Let $\mathbf{x}_{it}' = (1, \tilde{\mathbf{x}}_{it})$, where $\tilde{\mathbf{x}}_{it}$ denotes the $(1 \times (K-1))$ time-varying explanatory variables. Let $\boldsymbol{\alpha}_i' = (\alpha_{1i}, \tilde{\boldsymbol{\alpha}}_i')$ and $\bar{\boldsymbol{\beta}}' = (\bar{\boldsymbol{\beta}}_1, \tilde{\bar{\boldsymbol{\beta}}})$ be the corresponding partitions, Rewrite (6.9.28) in the form

$$E\begin{pmatrix} \alpha_{1i} & | & \\ & | & \mathbf{x}_{it} \\ \tilde{\boldsymbol{\alpha}}_i & | & \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{b}}_1' \\ \tilde{B} \end{pmatrix} (\tilde{\mathbf{x}}_{it} - E\tilde{\mathbf{x}}_{it}). \tag{6.9.41}$$

Then

$$\begin{aligned} E(\mathbf{y}_i \mid X_i) = X_i\bar{\boldsymbol{\beta}} &+ \mathbf{e}(\mathbf{x}_i - E\bar{\tilde{\mathbf{x}}}_i)'\tilde{\mathbf{b}}_1^* \\ &+ \tilde{X}_i \otimes [\tilde{X}_i - E(\tilde{X}_i)]' \operatorname{vec}(\hat{B}'), \end{aligned} \tag{6.9.42}$$

where $\tilde{\mathbf{b}}_1^* = T\mathbf{b}_1$, $\tilde{X}_i$ is the $T \times (K-1)$ stacked $\tilde{\mathbf{x}}_{it}'$ and $\mathbf{e}$ is a $(T \times 1)$ vector of 1's.

Therefore a consistent estimator of $\bar{\boldsymbol{\beta}}' = (\bar{\boldsymbol{\beta}}_1, \tilde{\bar{\boldsymbol{\beta}}})$ can be obtained by the least-squares regression of

$$\begin{aligned} y_{it} = \bar{\beta}_1 + \tilde{\mathbf{x}}_{it}'\tilde{\bar{\boldsymbol{\beta}}} &+ (\bar{\tilde{\mathbf{x}}}_i - \bar{\tilde{\mathbf{x}}})'\tilde{\mathbf{b}}_1^* \\ &+ (\tilde{\mathbf{x}}_{it}' \otimes (\tilde{\mathbf{x}}_{it} - \bar{\tilde{\mathbf{x}}})') \operatorname{vec}(\tilde{B}') + v_{it}, \end{aligned} \quad \begin{matrix} i = 1, \ldots, N, \\ t = 1, \ldots, T. \end{matrix} \tag{6.9.43}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

6.9.3.4 *Semiparametric Estimates*

The application of group mean estimator (6.9.23) requires precise $T > K$. If $T \leq K$ (Hsiao, Li, Liang, and Xie 2012) suggest a semiparametric estimator if there exists a q-dimensional random variable $\mathbf{z}_i (q < T \leq K)$ such that conditional on $\mathbf{z}_i$, $\boldsymbol{\beta}_i$ and $\mathbf{x}_{it}$ are conditionally independent, $(\boldsymbol{\beta}_i \perp \mathbf{x}_{it} \mid \mathbf{z}_i)$, and $E[\mathbf{x}_{it}\mathbf{x}_{it}' \mid \mathbf{z}_i]$ is a full rank matrix. Then $E(\boldsymbol{\alpha}_i \mid \mathbf{x}_i, \mathbf{z}_i) = E(\boldsymbol{\alpha}_i \mid \mathbf{z}_i) \equiv \mathbf{g}(\mathbf{z}_i)$. The random variable $\mathbf{z}_i$ can contain components of $\mathbf{x}_i$; it could be a function of $\mathbf{x}_i$; say, the propensity score (Rosenbaum and Rubin 1983); or an instrument for $\mathbf{x}_{it}$ and $\boldsymbol{\beta}_i$ (e.g., Heckman and Vytlacil 2001, 2005, 2007; Heckman, Schmierer, and Urzua 2011); or simply the time series mean of the ith individual's $\mathbf{x}_{it}$, $\bar{\mathbf{x}}_i$ (Mundlak 1978a).

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\bar{\boldsymbol{\beta}} + \mathbf{x}_{it}'\mathbf{g}(\mathbf{z}_i) + \epsilon_{it} \\ &= \mathbf{x}_{it}'\boldsymbol{\theta}(\mathbf{z}_i) + \epsilon_{it}, \end{aligned} \tag{6.9.44}$$

where $\boldsymbol{\theta}(\mathbf{z}_i) = \bar{\boldsymbol{\beta}} + \mathbf{g}(\mathbf{z}_i)$ and $E\mathbf{g}(\mathbf{z}_i) = \mathbf{0}$. Given $\boldsymbol{\theta}(\mathbf{z}_i)$ and $E(\boldsymbol{\theta}(\mathbf{z}_i)) = \bar{\boldsymbol{\beta}}$, a consistent estimator of $\bar{\boldsymbol{\beta}}$ is

$$\hat{\bar{\boldsymbol{\beta}}}_{semi} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\theta}(\mathbf{z}_i). \tag{6.9.45}$$

Hsiao, Li, Liang, and Xie (2012) suggest two types of semiparametric estimators for $\boldsymbol{\theta}(\mathbf{z})$: local constant and local polynomial estimation methods. The local constant estimator of $\boldsymbol{\theta}(\mathbf{z})$ is given by

$$\hat{\boldsymbol{\theta}}_{LC}(\mathbf{z}) = \left(\sum_{j=1}^{N}\sum_{s=1}^{T} x_{js}x_{js}' K_{h,z_{jz}}\right)^{-1} \sum_{j=1}^{N}\sum_{s=1}^{T} x_{js}y_{js}K_{h,z_{jz}}, \tag{6.9.46}$$

where $K_{h,z_{jz}} = \Pi_{l=1}^{q} k(\frac{z_{jl}-z_l}{h_l})$ is the product kernel, $k(\cdot)$ is the univariate kernel function, and z_{jl} and z_l are the lth-component of $\mathbf{z}_j$ and $\mathbf{z}$, respectively.

The local polynomial estimation minimizes the kernel weighted sum of squared errors

$$\sum_{j=1}^{N}\sum_{s=1}^{T}\left[y_{js} - \sum_{0\leq|k|\leq p} x_{js}' b_k(\mathbf{z})(\mathbf{z}_j - \mathbf{z})^k\right]^2 K_{h,z_{jz}}, \tag{6.9.47}$$

with respect to each $b_k(\mathbf{z})$ which gives an estimate of $\hat{b}_k(\mathbf{z})$, and $k\hat{b}_k(\mathbf{z})$ estimates $D^k\boldsymbol{\theta}(\mathbf{z})$. Thus, $\hat{\boldsymbol{\theta}}_{LP} = \hat{\mathbf{b}}_0(\mathbf{z})$ is the pth order local polynomial estimator of $\boldsymbol{\theta}(\mathbf{z})$.

Simple Monte Carlo studies conducted by Hsiao et al. (2012) show that if the exact order of $E(\boldsymbol{\alpha}_i \mid \mathbf{x}_i)$ is known, the panel pooled least-squares estimator performs well. If $E(\boldsymbol{\alpha}_i \mid \mathbf{x}_i)$ is unknown, the group mean estimator ((6.9.23) or (6.9.45)) semiparametric estimator is robust to a variety of joint distribution of $(\boldsymbol{\alpha}_i, \mathbf{x}_{it})$, but not the conventional fixed-effects estimator.

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

APPENDIX 6A: COMBINATION OF TWO NORMAL DISTRIBUTIONS

Suppose that conditional on X, $\boldsymbol{\beta}$, $\mathbf{y} \sim N(X\boldsymbol{\beta}, \Omega)$ and $\boldsymbol{\beta} \sim N(A\bar{\boldsymbol{\beta}}, C)$. Then the posterior of $\boldsymbol{\beta}$ and $\bar{\boldsymbol{\beta}}$ given $\mathbf{y}$ is

$$P(\boldsymbol{\beta}, \bar{\boldsymbol{\beta}} \mid \mathbf{y}) \propto \exp\left[-\frac{1}{2}\{(\mathbf{y} - X\boldsymbol{\beta})'\Omega^{-1}(\mathbf{y} - X\boldsymbol{\beta}) + (\boldsymbol{\beta} - A\bar{\boldsymbol{\beta}})'C^{-1}(\boldsymbol{\beta} - A\bar{\boldsymbol{\beta}})\}\right], \tag{6A.1}$$

where "$\propto$" denotes "proportionality." Using the identity (e.g., Rao 1971, p. 33)

$$(D + BFB')^{-1} = D^{-1} - D^{-1}B(B'D^{-1}B + F^{-1})^{-1}B'D^{-1}. \tag{6A.2}$$

and

$$(D + F)^{-1} = D^{-1} - D^{-1}(D^{-1} + F^{-1})^{-1}D^{-1}, \tag{6A.3}$$

we can complete the squares of

$$\begin{aligned}(\boldsymbol{\beta} - A\bar{\boldsymbol{\beta}})'C^{-1}(\boldsymbol{\beta} - A\bar{\boldsymbol{\beta}}) + (\mathbf{y} - X\boldsymbol{\beta})'\Omega^{-1}(\mathbf{y} - X\boldsymbol{\beta}) \\ = \boldsymbol{\beta}'C^{-1}\boldsymbol{\beta} + \bar{\boldsymbol{\beta}}'A'C^{-1}A\bar{\boldsymbol{\beta}} - 2\boldsymbol{\beta}'C^{-1}A\bar{\boldsymbol{\beta}} \\ + \mathbf{y}'\Omega^{-1}\mathbf{y} + \boldsymbol{\beta}'X'\Omega^{-1}X\boldsymbol{\beta} - 2\boldsymbol{\beta}'X'\Omega^{-1}\mathbf{y}.\end{aligned} \tag{6A.4}$$

Let

$$\begin{aligned}Q_1 = [\boldsymbol{\beta} - (X'\Omega^{-1}X + C^{-1})^{-1}(X\Omega^{-1}\mathbf{y} + C^{-1}A\bar{\boldsymbol{\beta}})]'(C^{-1} + X'\Omega^{-1}X) \\ \cdot [\boldsymbol{\beta} - (X'\Omega^{-1}X + C^{-1})^{-1}(X'\Omega^{-1}\mathbf{y} + C^{-1}A\bar{\boldsymbol{\beta}})].\end{aligned} \tag{6A.5}$$

then

$$\begin{aligned}&\boldsymbol{\beta}'C^{-1}\boldsymbol{\beta} + \boldsymbol{\beta}'X'\Omega^{-1}X\boldsymbol{\beta} - 2\boldsymbol{\beta}'C^{-1}A\bar{\boldsymbol{\beta}} - 2\boldsymbol{\beta}'X'\Omega^{-1}\mathbf{y} \\ &= Q_1 - (X'\Omega^{-1}\mathbf{y} + C^{-1}A\bar{\boldsymbol{\beta}})'(X'\Omega^{-1}X + C^{-1})^{-1}(X'\Omega^{-1}\mathbf{y} + C^{-1}A\bar{\boldsymbol{\beta}}).\end{aligned} \tag{6A.6}$$

Substituting (6A.6) into (6A.4) yields

$$\begin{aligned}&Q_1 + \mathbf{y}'[\Omega^{-1} - \Omega^{-1}X(X'\Omega^{-1}X + C^{-1})^{-1}X'\Omega^{-1}]\mathbf{y} \\ &\quad + \bar{\boldsymbol{\beta}}'A'[C^{-1} - C^{-1}(X'\Omega^{-1}X + C^{-1})^{-1}C^{-1}]A\bar{\boldsymbol{\beta}} \\ &\quad - 2\bar{\boldsymbol{\beta}}'A'C^{-1}(X'\Omega^{-1}X + C^{-1})^{-1}X'\Omega^{-1}\mathbf{y} \\ &\quad = Q_1 + \mathbf{y}'(XCX' + \Omega)^{-1}\mathbf{y} + \bar{\boldsymbol{\beta}}'A'X'(XCX' + \Omega)^{-1}XA\bar{\boldsymbol{\beta}} \\ &\quad\quad - 2\bar{\boldsymbol{\beta}}'A'X'(XCX' + \Omega)^{-1}\mathbf{y} \\ &\quad = Q_1 + Q_2 + Q_3\end{aligned} \tag{6A.7}$$

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press

where

$$Q_2 = \{\bar{\boldsymbol{\beta}} - [A'X'(XCX' + \Omega)^{-1}XA]^{-1}[A'X'(XCX' + \Omega)^{-1}y]\}' \cdot [A'X'(XCX' + \Omega)^{-1}XA] \cdot \{\bar{\boldsymbol{\beta}} - [A'X'(XCX' + \Omega)^{-1}XA]^{-1}[A'X'(XCX' + \Omega)^{-1}\mathbf{y}]\}, \tag{6A.8}$$

$$Q_3 = \mathbf{y}'\{(XCX' + \Omega)^{-1} - (XCX' + \Omega)^{-1}XA[A'X(XCX' + \Omega)^{-1}XA]^{-1} \cdot A'X'(XCX' + \Omega)^{-1}\}\ \mathbf{y} \tag{6A.9}$$

Since Q_3 is a constant independent of $\boldsymbol{\beta}$ and $\bar{\boldsymbol{\beta}}$, we can write $P(\boldsymbol{\beta}, \bar{\boldsymbol{\beta}} \mid \mathbf{y})$ in the form of $P(\boldsymbol{\beta} \mid \bar{\boldsymbol{\beta}}, \mathbf{y})P(\bar{\boldsymbol{\beta}} \mid \mathbf{y})$, which becomes

$$P\{\boldsymbol{\beta}, \bar{\boldsymbol{\beta}} \mid \mathbf{y}\} \propto \exp\left\{-\frac{1}{2}Q_1\right\} \exp\left\{-\frac{1}{2}Q_2\right\} \tag{6A.10}$$

where $\exp\{-\frac{1}{2}Q_1\}$ is proportional to $P(\boldsymbol{\beta} \mid \bar{\boldsymbol{\beta}}, \mathbf{y})$ and $\exp\{-\frac{1}{2}Q_2\}$ is proportional to $P(\bar{\boldsymbol{\beta}} \mid \mathbf{y})$. That is, $P(\boldsymbol{\beta} \mid \bar{\boldsymbol{\beta}}, \mathbf{y})$ is $N\{(X'\Omega^{-1}X + C^{-1})^{-1}(X'\Omega^{-1}\mathbf{y} + C^{-1}A\bar{\boldsymbol{\beta}}), (C^{-1} + X'\Omega^{-1}X)^{-1}\}$ and $P(\bar{\boldsymbol{\beta}} \mid \mathbf{y})$ is $N\{[A'X'(XCX' + \Omega)^{-1}XA]^{-1}[A'X'(XCX' + \Omega)^{-1}\mathbf{y}]^{-1}, [A'X'(XCX' + \Omega)^{-1}XA]^{-1}\}$.

Alternatively, we may complete the square of the left side of (6A.4) with the aim of writing $P(\boldsymbol{\beta}, \bar{\boldsymbol{\beta}} \mid \mathbf{y})$ in the form of $P(\bar{\boldsymbol{\beta}} \mid \boldsymbol{\beta}, \mathbf{y})P(\boldsymbol{\beta} \mid \mathbf{y})$,

$$\begin{aligned} &Q_4 + \boldsymbol{\beta}'[X'\Omega^{-1}X + C^{-1} - C^{-1}A(A'CA)^{-1}A'C^{-1}]\boldsymbol{\beta} \\ &\quad - 2\boldsymbol{\beta}'X'\Omega^{-1}\mathbf{y} + \mathbf{y}'\Omega^{-1}\mathbf{y} \\ &\quad = Q_4 + Q_5 + Q_3, \end{aligned} \tag{6A.11}$$

where

$$Q_4 = [\bar{\boldsymbol{\beta}} - (A'C^{-1}A)^{-1}A'C^{-1}\boldsymbol{\beta}]'(A'C^{-1}A) \cdot [\bar{\boldsymbol{\beta}} - (A'C^{-1}A)^{-1}A'C^{-1}\boldsymbol{\beta}], \tag{6A.12}$$

$$Q_5 = [\boldsymbol{\beta} - D^{-1}X'\Omega^{-1}\mathbf{y}]'D[\boldsymbol{\beta} - D^{-1}X'\Omega^{-1}\mathbf{y}]. \tag{6A.13}$$

and

$$D = X'\Omega^{-1}X + C^{-1} - C^{-1}A(A'C^{-1}A)^{-1}A'C^{-1}. \tag{6A.14}$$

Therefore, $P(\bar{\boldsymbol{\beta}} \mid \boldsymbol{\beta}, \mathbf{y}) \sim N\{(A'C^{-1}A)^{-1}C^{-1}\boldsymbol{\beta}, (A'C^{-1}A)^{-1}\}$ and $P(\boldsymbol{\beta} \mid \mathbf{y}) \sim N\{D^{-1}X'\Omega^{-1}\ \mathbf{y}, D^{-1}\}$.

https://doi.org/10.1017/CBO9781139839327.009 Published online by Cambridge University Press