larger than T. Note what happens if $n > T$. In this case, the $n \times n$ matrix has rank T, which is less than n, so it must be singular, and the FGLS estimator cannot be computed. Consider Example 10.2. We aggregated the 48 states into $n = 9$ regions. It would not be possible to fit a full model for the $n = 48$ states with only $T = 17$ observations. This result is a deficiency of the data set, not the model. The population matrix, Σ is positive definite. But, if there are not enough observations, then the data set is too short to obtain a positive definite estimate of the matrix.

11.10 Systems of EQUATIONS ~~10.3 PANEL DATA APPLICATIONS~~

Extensions of the SUR model to panel data applications have been made in two directions. Several studies have layered the familiar random effects treatment of Section ~~9.5~~ 11.5 on top of the generalized regression. An alternative treatment of the fixed and random effects models as a form of seemingly unrelated regressions model suggested by Chamberlain (1982, 1984) has provided some of the foundation of recent treatments of dynamic panel data models, as in Sections 11.8.2 and 11.8.3.

11.10.1 ~~10.3.1~~ RANDOM EFFECTS SUR MODELS

Avery (1977) suggested a natural extension of the random effects model to multiple equations,

$$y_{it,j} = \mathbf{x}'_{it,j}\boldsymbol{\beta}_j + \varepsilon_{it,j} + u_{i,j},$$

where j indexes the equation, i indexes individuals, and t is the time index as before. Each equation can be treated as a random effects model. In this instance, however, the efficient estimator when the equations are *actually* unrelated (that is, $\text{Cov}[\varepsilon_{it,m}, \varepsilon_{it,l} \,|\, \mathbf{X}] = 0$ and $\text{Cov}[u_{i,m}, u_{i,l} \,|\, \mathbf{X}] = 0$) is equation by equation GLS as developed in Section ~~9.5,~~ 11.5 not OLS. That is, without the cross-equation correlation, each equation constitutes a random effects model. The cross-equation correlation takes the form

$$E[\varepsilon_{it,j}\varepsilon_{it,l} \,|\, \mathbf{X}] = \sigma_{jl}$$

and

$$E[u_{i,j}u_{i,l} \,|\, \mathbf{X}] = \theta_{jl}.$$

Observations remain uncorrelated across individuals, $(\varepsilon_{it,j}, \varepsilon_{rs,l})$ and $(u_{ij}, u_{r,l})$ when $i \neq r$. The "noise" terms, $\varepsilon_{it,j}$ are also uncorrelated across time for all individuals and across individuals. Correlation over time arises through the influence of the common effect, which produces persistent random effects for the given individual, both within the equation and across equations through θ_{jl}. Avery developed a two-step estimator for the model. At the first step, as usual, estimates of the variance components are based on OLS residuals. The second step is FGLS. Subsequent studies have added features to the model. Magnus (1982) derived the log likelihood function for normally distributed disturbances, the likelihood equations for the MLE, and a method of estimation. Verbon (1980) added heteroscedasticity to the model.

There have also been a handful of applications, including Howrey and Varian's (1984) analysis of electricity pricing and the impact of time of day rates, Brown et al.'s

11-89

(1983) treatment of a form and of the capital asset pricing model (CAPM), Sickles's (1985) analysis of airline costs, ~~and~~ Wan et al.'s (1992) development of a nonlinear panel data SUR model for agricultural output, ~~and~~.

11.18

Example ~~10.5~~ 11.18 Demand for Electricity and Gas

Beierlein, Dunn, and McConnon (1981) proposed a dynamic panel data SUR model for demand for electricity and natural gas in the northeastern U.S. United States. The central equation of the model is

$$\begin{aligned}\ln Q_{it,j} &= \beta_0 + \beta_1 \ln P\_natural\ gas_{it,j} + \beta_2 \ln P\_electricity_{it,j} + \beta_3 \ln P\_fuel\ oil_{it,j}\\ &\quad + \beta_4 \ln per\ capita\ income_{it,j} + \beta_5 \ln Q_{i,t-1,j} + w_{it,j}\\ w_{it,j} &= \varepsilon_{it,j} + u_{i,j} + v_{t,j}\end{aligned}$$

where

j = consuming sectors (natural gas, electricity) × (residential, comercial, industrial)

i = state (New England plus New York, New Jersey, Pennsylvania)

t = year, 1957, ..., 1977.

Note that this model has both time and state random effects and a lagged dependent variable in each equation.

10.3.2 THE RANDOM AND FIXED EFFECTS MODELS

The linear unobserved effects model is

$$y_{it} = c_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}. \tag{10-22}$$

The **random effects** model assumes that $E[c_i \,|\, \mathbf{X}_i] = \alpha$, where the T rows of $\mathbf{X}_i$ are $\mathbf{x}_{it}'$. As we saw in Section 9.5, this model can be estimated consistently by ordinary least squares. Regardless of how ε_{it} is modeled, there is autocorrelation induced by the common, unobserved c_i, so the generalized regression model applies. The random effects formulation is based on the assumption $E[\mathbf{w}_i\mathbf{w}_i' \,|\, \mathbf{X}_i] = \sigma_\varepsilon^2\mathbf{I}_T + \sigma_u^2\mathbf{ii}'$, where $w_{it} = (\varepsilon_{it} + u_i)$. We developed the GLS and FGLS estimators for this formulation as well as a strategy for robust estimation of the OLS covariance matrix. Among the implications of the development of Section 10.2.8 is that this formulation of the disturbance covariance matrix is more restrictive than necessary, given the information contained in the data. The assumption that $E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' \,|\, \mathbf{X}_i] = \sigma_\varepsilon^2\mathbf{I}_T$ assumes that the correlation across periods is equal for all pairs of observations, and arises solely through the persistent c_i. In Section 10.2.8, we estimated the equivalent model with an unrestricted covariance matrix, $E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' \,|\, \mathbf{X}_i] = \boldsymbol{\Sigma}$. The implication is that the random effects treatment includes two restrictive assumptions, mean independence, $E[c_i \,|\, \mathbf{X}_i] = \alpha$, and homoscedasticity, $E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' \,|\, \mathbf{X}_i] = \sigma_\varepsilon^2\mathbf{I}_T$. [We do note, dropping the second assumption will cost us the identification of σ_u^2 as an estimable parameter. This makes sense—if the correlation across periods t and s can arise from either their common u_i or from correlation of $(\varepsilon_{it}, \varepsilon_{is})$ then there is no way for us separately to estimate a variance for u_i apart from the covariances of ε_{it} and ε_{is}.] It is useful to note, however, that the panel data model can be viewed and formulated as a seemingly unrelated regressions model with common coefficients in which each period constitutes an equation, Indeed, it is possible, albeit unnecessary, to impose the restriction $E[\mathbf{w}_i\mathbf{w}_i' \,|\, \mathbf{X}_i] = \sigma_\varepsilon^2\mathbf{I}_T + \sigma_u^2\mathbf{ii}'$.

11-90

For each local authority, this model implies

$$y_{it} = \gamma_t + \mathbf{x}_{it}'\boldsymbol{\beta} + u_i + \lambda\Sigma_j w_{ij}\varepsilon_{jt} + v_{it}.$$

The authors use maximum likelihood to estimate the parameters of the model. To simplify the computations, they note that the maximization can be done using a two-step procedure. As we have seen in other applications, when $\boldsymbol{\Omega}$ in a generalized regression model is known, the appropriate estimator is GLS. For both of these models, with known spatial autocorrelation parameter, a GLS transformation of the data produces a classical regression model. [See (8-11).] The method used is to iterate back and forth between simple OLS estimation of γ_t, $\boldsymbol{\beta}$ and σ_ε^2 and maximization of the "concentrated log likelihood" function which, given the other estimates, is a function of the spatial autocorrelation parameter, ρ or λ, and the variance of the heterogeneity, σ_u^2.

The dependent variable in the models is the log of per capita mental health expenditures. The covariates are the percentage of males and of people under 20 in the area, average mortgage rates, numbers of unemployment claims, employment, average house price, median weekly wage, percent of single parent households, dummy variables for Labour party or Liberal Democrat party authorities, and the density of population ("to control for supply-side factors"). The estimated spatial autocorrelation coefficients for the two models are 0.1579 and 0.1220, both more than twice as large as the estimated standard error. Based on the simple Wald tests, the hypothesis of no spatial correlation would be rejected. The log likelihood values for the two spatial models were +206.3 and +202.8, compared to −211.1 for the model with no spatial effects or region effects, so the results seem to favor the spatial models based on a chi-squared test statistic (with one degree of freedom) of twice the difference. However, there is an ambiguity in this result as the improved "fit" could be due to the region effects rather than the spatial effects. A simple random effects model shows a log likelihood value of +202.3, which bears this out. Measured against this value, the spatial lag model seems the preferred specification, whereas the spatial autocorrelation model does not add significantly to the log likelihood function compared to the basic random effects model.

11.11 ~~9.8~~ PARAMETER HETEROGENEITY

The treatment so far has essentially treated the slope parameters of the model as fixed constants, and the intercept as varying randomly from group to group. An equivalent formulation of the pooled, fixed, and random effects model is

$$y_{it} = (\alpha + u_i) + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it},$$

where u_i is a person-specific random variable with conditional variance zero in the pooled model, positive in the others, and conditional mean dependent on $\mathbf{X}_i$ in the fixed effects model and constant in the random effects model. By any of these, the heterogeneity in the model shows up as variation in the constant terms in the regression model. There is ample evidence in many studies—we will examine two later—that suggests that the other parameters in the model also vary across individuals. In the dynamic model we consider in Section 11.11.3 ~~9.8.5~~, cross-country variation in the slope parameter in a production function is the central focus of the analysis. This section will consider several approaches to analyzing parameter heterogeneity in panel data models. ~~The model will be extended to multiple equations in Section 10.3.~~

11.11.1 ~~9.8.1~~ THE RANDOM COEFFICIENTS MODEL

Parameter heterogeneity across individuals or groups can ~~me~~ be modeled as stochastic variation.[26] Suppose that we write

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i,$$
$$E[\boldsymbol{\varepsilon}_i \mid \mathbf{X}_i] = \mathbf{0},$$
$$E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' \mid \mathbf{X}_i] = \sigma_\varepsilon^2\mathbf{I}_T, \qquad \text{11-82 } \cancel{(9\text{-}47)}$$

where

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{u}_i \qquad \text{11-83 } \cancel{(9\text{-}48)}$$

and

$$E[\mathbf{u}_i \mid \mathbf{X}_i] = \mathbf{0},$$
$$E[\mathbf{u}_i\mathbf{u}_i' \mid \mathbf{X}_i] = \boldsymbol{\Gamma}. \qquad \text{11-84 } \cancel{(9\text{-}49)}$$

(Note that if only the constant term in $\boldsymbol{\beta}$ is random in this fashion and the other parameters are fixed as before, then this reproduces the random effects model we studied in Section ~~9.5~~ 11.5.) Assume for now that there is no autocorrelation or cross-section correlation in $\boldsymbol{\varepsilon}_i$. We also assume for now that $T > K$, so that when desired, it is possible to compute the linear regression of $\mathbf{y}_i$ on $\mathbf{X}_i$ for each group. Thus, the $\boldsymbol{\beta}_i$ that applies to a particular cross-sectional unit is the outcome of a random process with mean vector $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Gamma}$.[27] By inserting ~~(9-48)~~ (11-83) into ~~(9-47)~~ (11-82) and expanding the result, we obtain a generalized regression model for each block of observations:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + (\boldsymbol{\varepsilon}_i + \mathbf{X}_i\mathbf{u}_i),$$

so

$$\boldsymbol{\Omega}_{ii} = E[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})' \mid \mathbf{X}_i] = \sigma_\varepsilon^2\mathbf{I}_T + \mathbf{X}_i\boldsymbol{\Gamma}\mathbf{X}_i'.$$

For the system as a whole, the disturbance covariance matrix is block diagonal, with $T \times T$ diagonal block $\boldsymbol{\Omega}_{ii}$. We can write the GLS estimator as a matrix weighted average of the group specific OLS estimators:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} = \sum_{i=1}^{n}\mathbf{W}_i\mathbf{b}_i, \qquad \text{11-85 } \cancel{(9\text{-}50)}$$

where

$$\mathbf{W}_i = \left[\sum_{i=1}^{n}\left(\boldsymbol{\Gamma} + \sigma_\varepsilon^2(\mathbf{X}_i'\mathbf{X}_i)^{-1}\right)^{-1}\right]^{-1}\left(\boldsymbol{\Gamma} + \sigma_\varepsilon^2(\mathbf{X}_i'\mathbf{X}_i)^{-1}\right)^{-1}.$$

27 [26] The most widely cited studies are Hildreth and Houck (1968), Swamy (1970, 1971, 1974), Hsiao (1975), and Chow (1984). See also Breusch and Pagan (1979). Some recent discussions are Swamy and Tavlas (1995, 2001) and Hsiao (2003). The model bears some resemblance to the Bayesian approach of Chapter 18. But, the similarity is only superficial. We are maintaining the classical approach to estimation throughout.

28 [27] Swamy and Tavlas (2001) label this the "first-generation random coefficients model" (RCM). We will examine the "second generation" (the current generation) of random coefficients models in the next section.

11-92

Empirical implementation of this model requires an estimator of $\mathbf{\Gamma}$. One approach [see, e.g., Swamy (1971)] is to use the empirical variance of the set of n least squares estimates, $\mathbf{b}_i$ minus the average value of $s_i^2(\mathbf{X}_i'\mathbf{X}_i)^{-1}$:

$$\mathbf{G} = [1/(n-1)]\left[\Sigma_i \mathbf{b}_i\mathbf{b}_i' - n\bar{\mathbf{b}}\bar{\mathbf{b}}'\right] - (1/N)\Sigma_i \mathbf{V}_i, \qquad \text{88 / 11-86 / ~~(9-51)~~}$$

where

$$\bar{\mathbf{b}} = (1/n)\Sigma_i \mathbf{b}_i$$

and

$$\mathbf{V}_i = s_i^2(\mathbf{X}_i'\mathbf{X}_i)^{-1}.$$

This matrix may not be positive definite, however, in which case [as Baltagi (2005) suggests], one might drop the second term.

A chi-squared test of the random coefficients model against the alternative of the classical regression (no randomness of the coefficients) can be based on

$$C = \Sigma_i(\mathbf{b}_i - \mathbf{b}_*)'\mathbf{V}_i^{-1}(\mathbf{b}_i - \mathbf{b}_*),$$

where

$$\mathbf{b}_* = \left[\Sigma_i \mathbf{V}_i^{-1}\right]^{-1}\Sigma_i \mathbf{V}_i^{-1}\mathbf{b}_i.$$

Under the null hypothesis of homogeneity, C has a limiting chi-squared distribution with $(n-1)K$ degrees of freedom. The best linear unbiased individual predictors of the group-specific coefficient vectors are matrix weighted averages of the GLS estimator, $\hat{\boldsymbol\beta}$, and the group specific OLS estimates, $\mathbf{b}_i$,[~~28~~ 29] FN 29

$$\hat{\boldsymbol\beta}_i = \mathbf{Q}_i\hat{\boldsymbol\beta} + [\mathbf{I} - \mathbf{Q}_i]\mathbf{b}_i, \qquad \text{89 / 11-87 / ~~(9-52)~~}$$

where

$$\mathbf{Q}_i = [(1/s_i^2)\mathbf{X}_i'\mathbf{X}_i + \mathbf{G}^{-1}]^{-1}\mathbf{G}^{-1}.$$

Example ~~9.12~~ 11.19 Random Coefficients Model

In Example ~~9.9~~ 10.1, we examined Munell's production model for gross state product,

$$\ln gsp_{it} = \beta_1 + \beta_2 \ln \text{~~p\_cap~~ pc}_{it} + \beta_3 \ln hwy_{it} + \beta_4 \ln water_{it}$$
$$+ \beta_5 \ln util_{it} + \beta_6 \ln emp_{it} + \beta_7\, unemp_{it} + \varepsilon_{it}, \quad i = 1, \ldots, 48;\ t = 1, \ldots, 17.$$

The panel consists of state level data for 17 years. The model in Example ~~9.9~~ 10.1 (and Munnell's) provide no means for parameter heterogeneity save for the constant term. We have reestimated the model using the Hildreth and Houck approach. The OLS~~,~~ and Feasible GLS ~~and maximum likelihood estimates~~ are given in Table ~~9.5~~ 11.14. The chi-squared statistic for testing the null hypothesis of parameter homogeneity is 25,556.26, with 7(47) = 329 degrees of freedom. The critical value from the table is 372.299, so the hypothesis would be rejected.

TB 11.14

Unlike the other cases we have examined in this chapter, the FGLS estimates are very different from OLS in these estimates, in spite of the fact that both estimators are consistent and the sample is fairly large. The underlying standard deviations are computed using $\mathbf{G}$ as the covariance matrix. [For these data, subtracting the second matrix rendered $\mathbf{G}$ not positive

[~~28~~ 29] See Hsiao (2003, pp. 144–149).

11-93

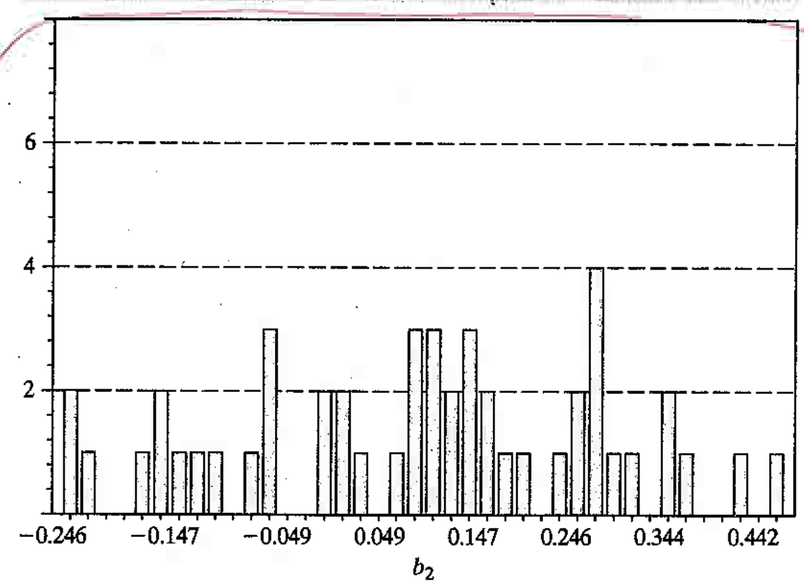
TABLE ~~9.9~~ 11.14 Estimated Random Coefficients Models

	Least Squares		Feasible GLS			Maximum Simulated Likelihood	
Variable	*Estimate*	*Standard Error*	*Estimate*	*Standard Error*	*Popn. Std. Deviation*	*Estimate*	*Std. Error*
Constant	1.9260	0.05250	1.6533	1.08331	7.0782	1.9463 (0.0411)	0.03569
ln pc	0.3120	0.01109	0.09409	0.05152	0.3036	0.2962 (0.0730)	0.00882
ln hwy	0.05888	0.01541	0.1050	0.1736	1.1112	0.09515 (0.146)	0.01157
ln water	0.1186	0.01236	0.07672	0.06743	0.4340	0.2434 (0.343)	0.01929
ln util	0.00856	0.01235	−0.01489	0.09886	0.6322	−0.1855 (0.281)	0.02713
ln emp	0.5497	0.01554	0.9190	0.1044	0.6595	0.6795 (0.121)	0.02274
unemp	−0.00727	0.001384	−0.004706	0.002067	0.01266	−0.02318 (0.0308)	0.002712
σ_ε	0.08542		0.2129			0.02748	
ln L	853.1372					1567.233	

11-86 definite, so in the table, the standard deviations are based on the estimates using only the first term in (9-51).] The increase in the standard errors is striking. This suggests that there is considerable variation in the parameters across states. We have used (9-52) to compute the estimates of the state specific coefficients. Figure ~~9.1~~ 11.1 shows a histogram for the coefficient on private capital. As suggested, there is a wide variation in the estimates. 11-87

FIG 11.1

FIGURE ~~9.1~~ 11.1 Estimates of Coefficients on Private Capital.



11.11.2 A Hierarchical Linear Model

Many researchers have employed a two-step approach to estimate two level models. In a common form of the application, a panel data set is employed to estimate the model,

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it}\,,\; i = 1, \ldots, n,\; t = 1, \ldots, T,$$

$$\beta_{i,k} = \mathbf{z}_i'\boldsymbol{\alpha}_k + u_{i,k},\; i = 1, \ldots, n.$$

FN 30 — and

Assuming the panel is long enough, the first equation is estimated n times, once for each individual i, then the estimated coefficient on x_{itk} in each regression forms an observation for the second step regression.[32] 30

minus (This is the approach we took in the previous section; each a_i is computed by a linear regression of y_i - $\mathbf{X}_i\mathbf{b}_{LSDV}$ on a column of ones.)

Au: Is sub "LSDV" OK italic?

30 [32] An extension of the model in which u_i is heteroscedastic is developed at length in Saxonhouse (1976) and revisited by Achen (2005).

Example ~~9.15~~ 11.20 Fannie Mae's Pass Through

Fannie Mae is the popular name for the Federal National Mortgage Corporation. Fannie Mae is the secondary provider for mortgage money for nearly all of the small and moderate sized home mortgages in the United States. Loans in the study described here are termed "small" if they are for less than $100,000. A loan is termed a "conforming" in the language of the literature on this market if (as of 2004), it was for no more than $333,700. A larger than conforming loan is called a "jumbo" mortgage. Fannie Mae provides the capital for nearly all conforming loans and no nonconforming loans. The question pursued in the study described here was whether the clearly observable spread between the rates on jumbo loans and conforming loans reflects the cost of raising the capital in the market. Fannie Mae is a "Government Sponsored Enterprice" (GSE). It was created by the U.S. Congress, but it is not an arm of the government; it is a private corporation. In spite of, or perhaps because of this ambiguous relationship to the government, apparently, capital markets believe that there is some benefit to Fannie Mae in raising capital. Purchasers of the GSE's debt securities seem to believe that the debt is implicitly backed by the government — this in spite of the fact that Fannie Mae explicitly states otherwise in its publications. This emerges as a "funding advantage" (GFA) estimated by the authors of the study of about 17 basis points (hundredths of one percent). In a study of the residential mortgage market, Passmore (2005) and Passmore, Sherlund, and Burgess (2005) sought to determine whether this implicit subsity to the GSE was passed on to the mortgagees or was, instead, passed on to the stockholders. Their approach utilitized a very large data set and a two-level, two-step estimation procedure. The first step equation estimated was a mortgage rate equation using a sample of roughly 1 million closed mortgages. All were conventional 30-year fixed-rate loans closed between April 1997 and May 2003. The dependent variable of interest is the rate on the mortgage, RM_{it}. The first level equation is subsidy

$$RM_{it} = \beta_{1i} + \beta_{2,i}\,J_{it} + \text{terms for “loan to value ratio,” “new home dummy variable,” “small mortgage”} + \text{terms for “fees charged” and whether the mortgage was originated by a mortgage company} + \varepsilon_{it}.$$

The main variable of interest in this model is J_{it}, which is a dummy variable for whether the loan is a jumbo mortgage. The "i" in this setting is a (state, time) pair for California, New Jersey, Maryland, Virginia, and all other states, and months from April 1997 to May 2003. There were 370 groups in total. The regression model was estimated for each group. At the second step, the coefficient of interest is $\beta_{2,i}$. On overall average, the spread between jumbo and conforming loans at the time was roughly 16 basis points. The second level equation is

11-95

$\beta_{2,i} = \alpha_1 + \alpha_2 \text{GFA}_i$
+ α_3 one-year treasury rate
+ α_4 ten-year treasure rate
+ α_5 credit risk
+ α_6 prepayment risk
+ measures of maturity mismatch risk
+ quarter and state fixed effects
+ mortgage market capacity
+ mortgage market development
+ u_i

The result ultimately of interest is the coefficient on GFA, α_2, which is interpreted as the fraction of the GSE funding advantage that is passed through to the mortgage holders. Four different estimates of α_2 were obtained, based on four different measures of corporate debt liquidity; the estimated values were $\left(\hat{\alpha}_2^1, \hat{\alpha}_2^2, \hat{\alpha}_2^3, \hat{\alpha}_2^4\right) = (0.07,\ 0.31,\ 0.17,\ 0.10)$. The four estimates were averaged using a **minimum distance estimator** (MDE). Let $\hat{\Omega}$ denote the estimated 4 × 4 asymptotic covariance matrix for the estimators. Denote the distance vector

$$\mathbf{d} = \left(\hat{\alpha}_2^1 - \alpha_2, \hat{\alpha}_2^2 - \alpha_2, \hat{\alpha}_2^3 - \alpha_2, \hat{\alpha}_2^4 - \alpha_2\right)'$$

The minimum distance estimator is the value for α_2 that minimizes $\mathbf{d}'\hat{\Omega}^{-1}\mathbf{d}$. For this study, $\hat{\Omega}$ is a diagonal matrix. It is straighforward to show that in this case, the MDE is

$$\hat{\alpha}_2 = \sum_{j=1}^4 \hat{\alpha}_2^j \left(\frac{1/\hat{\omega}_j}{\Sigma_{m=1}^4 1/\hat{\omega}_m} \right).$$

percent percent

The final answer is roughly 16%. By implication, then, the authors estimated that 84% of the GSE funding advantage was kept within the company or passed through to stockholders.

11-96

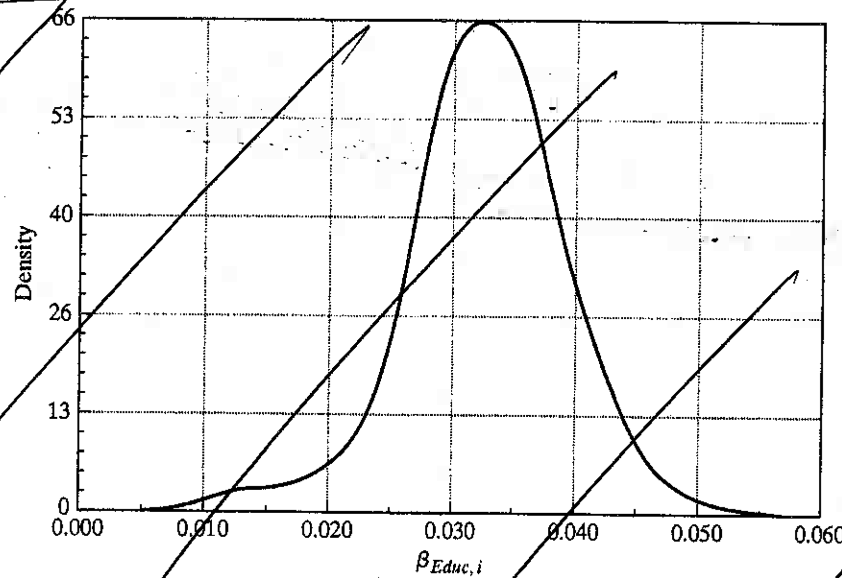


FIGURE 9.4 Kernel Density Estimate for Education Coefficient.

model is $\beta_{2,i} = \alpha_{2,1} + \alpha_{2,2}$ *Ability* $+ \beta_{2,3}$ *Mother's education* $+ \beta_{2,4}$ *Father's education* $+ u_{2,i}$. A rough indication of the magnitude of this result can be seen by inserting the sample means for these variables, 0.052374, 11.4719, and 11.7092, respectively. With these values, the mean value for the education coefficient is approximately 0.0327. This is comparable, though somewhat smaller, than the estimates for the pooled and random effects model. Of course, variation in this parameter across the sample individuals was the objective of this specification. Figure 9.4 plots a kernel density estimate for the 2,178 sample individuals. The figure shows the very wide range of variation in the sample estimates.

3 11.11.2 ~~9.8.5~~ PARAMETER HETEROGENEITY AND DYNAMIC PANEL DATA MODELS

The analysis ~~so far~~ in this section has involved static models and relatively straightforward estimation problems. We have seen as this section has progressed that parameter heterogeneity introduces a fair degree of complexity to the treatment. Dynamic effects in the model, with or without heterogeneity, also raise complex new issues in estimation and inference. There are numerous cases in which dynamic effects and parameter heterogeneity coincide in panel data models. This section will explore a few of the specifications and some applications. The familiar estimation techniques (OLS, FGLS, etc.) are not effective in these cases. The proposed solutions are developed in Chapter ~~12~~ 8 where we present the technique of instrumental variables and in Chapter ~~15~~ 13 where we present the GMM estimator and its application to **dynamic panel data models.**

AU: Term was already a KT in chapter. Here also?

11.21 *Example* ~~9.18~~ *Dynamic Panel Data Models*

The antecedent of much of the current research on panel data is Balestra and Nerlove's (1966) study of the natural gas market. [See, also, Nerlove (2002, Chapter 2).] The model is a stock-flow description of the derived demand for fuel for gas using appliances. The central equation is a model for total demand,

$$G_{it} = G^*_{it} + (1-r)G_{i,t-1},$$

11-97

where G_{it} is current total demand. Current demand consists of new demand, G^*_{it}, that is created by additions to the stock of appliances plus old demand which is a proportion of the previous period's demand, r being the depreciation rate for gas using appliances. New demand is due to net increases in the stock of gas using appliances, which is modeled as

$$G^*_{it} = \beta_0 + \beta_1 Price_{it} + \beta_2 \Delta Pop_{it} + \beta_3 Pop_{it} + \beta_4 \Delta Income_{it} + \beta_5 Income_{it} + \varepsilon_{it},$$

where Δ is the first difference (change) operator, $\Delta X_t = X_t - X_{t-1}$. The reduced form of the model is a dynamic equation,

$$G_{it} = \beta_0 + \beta_1 Price_{it} + \beta_2 \Delta Pop_{it} + \beta_3 Pop_{it} + \beta_4 \Delta Income_{it} + \beta_5 Income_{it} + \gamma G_{i,t-1} + \varepsilon_{it}.$$

The authors analyzed a panel of 36 states over a six-year period (1957–1962). Both fixed effects and random effects approaches were considered.

An equilibrium model for steady state growth has been used by numerous authors [e.g., Robertson and Symons (1992), Pesaran and Smith (1995), Lee, Pesaran, and Smith (1997), Pesaran, Shin, and Smith (1999), Nerlove (2002) and Hsiao, Pesaran, and Tahmiscioglu (2002)] for cross industry or country comparisons. Robertson and Symons modeled real wages in 13 OECD countries over the period 1958 to 1986 with a wage equation

$$W_{it} = \alpha_i + \beta_{1i} k_{it} + \beta_{2i} \Delta wedge_{it} + \gamma_i W_{i,t-1} + \varepsilon_{it},$$

where W_{it} is the real product wage for country i in year t, k_{it} is the capital-labor ratio, and *wedge* is the "tax and import price wedge."

Lee, Pesaran, and Smith (1997) compared income growth across countries with a steady-state income growth model of the form

$$\ln y_{it} = \alpha_i + \theta_i t + \lambda_i \ln y_{i,t-1} + \varepsilon_{it},$$

where $\theta_i = (1 - \lambda_i)\delta_i$, δ_i is the technological growth rate for country i and λ_i is the convergence parameter. The rate of convergence to a steady state is $1 - \lambda_i$.

Pesaran and Smith (1995) analyzed employment in a panel of 38 UK industries observed over 29 years, 1956–1984. The main estimating equation was

$$\ln e_{it} = \alpha_i + \beta_{1i} t + \beta_{2i} \ln y_{it} + \beta_{3i} \ln y_{i,t-1} + \beta_{4i} \ln \overline{y}_t + \beta_{5i} \ln \overline{y}_{t-1} + \beta_{6i} \ln w_{it} + \beta_{7i} \ln w_{i,t-1} + \gamma_{1i} \ln e_{i,t-1} + \gamma_{2i} \ln e_{i,t-2} + \varepsilon_{it},$$

where y_{it} is industry output, $\overline{y}_t$ is total (not average) output, and w_{it} is real wages.

KT KT KT

In the growth models, a quantity of interest is the **long run multiplier** or **long-run elasticity.** Long-run effects are derived through the following conceptual experiment. The essential feature of the models above is a dynamic equation of the form

$$y_t = \alpha + \beta x_t + \gamma y_{t-1}.$$

Suppose at time t, x_t is fixed from that point forward at $\overline{x}$. The value of y_t at that time will then be $\alpha + \beta\overline{x} + \gamma y_{t-1}$, given the previous value. If this process continues, and if $|\gamma| < 1$, then eventually y_s will reach an equilibrium at a value such that $y_s = y_{s-1} = \overline{y}$. If so, then $\overline{y} = \alpha + \beta\overline{x} + \gamma\overline{y}$, from which we can deduce that $\overline{y} = (\alpha + \overline{x})/(1-\gamma)$. ~~(We will analyze this computation at length in Chapter 13.)~~ The path to this equilibrium from time t into the future is governed by the **adjustment equation** KT

$$y_s - \overline{y} = (y_t - \overline{y})\gamma^{s-t}, s \geq t.$$

KT KT

The experiment, then, is to ask: What is the impact on the equilibrium of a change in the input, $\overline{x}$? The result is $\partial\overline{y}/\partial\overline{x} = \beta/(1-\gamma)$. This is the long-run multiplier, or **equilibrium multiplier** in the model. In the Pesaran and Smith model preceding, the inputs are in logarithms, so the multipliers are long-run elasticities. For example, with two lags of

11-98

In e_{it} in Pesaran and Smith's model, the long-run effects for wages are

$$\phi_i = (\beta_{6i} + \beta_{7i})/(1 - \gamma_{1i} - \gamma_{2i}).$$

In this setting, in contrast to the preceding treatments, the number of units, n, is generally taken to be fixed, though often it will be fairly large. The Penn World Tables (http://pwt.econ.upenn.edu/php_site/pwt_index.php) that provide the database for many of these analyses now contain information on almost 200 countries for well over 50 years. Asymptotic results for the estimators are with respect to increasing T, though we will consider in general, cases in which T is small. Surprisingly, increasing T and n at the same time need not simplify the derivations. We will revisit this issue in the next section.

URL

The parameter of interest in many studies is the average long-run effect, say $\overline{\phi} = (1/n)\Sigma_i\phi_i$, in the Pesaran and Smith example. Because n is taken to be fixed, the "parameter" $\overline{\phi}$ is a definable object of estimation—that is, with n fixed, we can speak of $\overline{\phi}$ as a parameter rather than as an estimator of a parameter. There are numerous approaches one might take. For estimation purposes, pooling, fixed effects, random effects, group means, or separate regressions are all possibilities. (Unfortunately, nearly all are inconsistent.) In addition, there is a choice to be made whether to compute the average of long-run effects or compute the long-run effect from averages of the parameters. The choice of the average of functions, $\overline{\phi}$ versus the function of averages,

$$\overline{\phi}* = \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_{6i} + \hat{\beta}_{7i})}{1 - \frac{1}{n}\sum_{i=1}^{n}(\hat{\gamma}_{1i} + \hat{\gamma}_{2i})}$$

turns out to be of substance. For their UK industry study, Pesaran and Smith report estimates of -0.33 for $\overline{\phi}$ and -0.45 for $\overline{\phi}*$. (The authors do not express a preference for one over the other.)

The development to this point is implicitly based on estimation of separate models for each unit (country, industry, etc.). There are also a variety of other estimation strategies one might consider. We will assume for the moment that the data series are stationary in the dimension of T. (See Chapter ~~22~~.) This is a transparently false assumption, as revealed by a simple look at the trends in macroeconomic data, but maintaining it for the moment allows us to proceed. We will reconsider it later. 23

We consider the generic, dynamic panel data model,

90

$$y_{it} = \alpha_i + \beta_i x_{it} + \gamma_i y_{i,t-1} + \varepsilon_{it}. \qquad \text{11-88} \ \cancel{(9\text{-}63)}$$

Assume that T is large enough that the individual regressions can be computed. In the absence of autocorrelation in ε_{it}, it has been shown [e.g., Griliches (1961), Maddala and Rao (1973)] that the OLS estimator of γ_i is biased downward, but consistent in T. Thus, $E[\hat{\gamma}_i - \gamma_i] = \theta_i/T$ for some θ_i. The implication for the individual estimator of the long-run multiplier, $\phi_i = \beta_i/(1-\gamma_i)$, is unclear in this case, however. The denominator is overestimated. But it is not clear whether the estimator of β_i is overestimated or underestimated. It is true that whatever bias there is $O(1/T)$. For this application, T is fixed and possibly quite small. The end result is that it is unlikely that the individual estimator of ϕ_i is unbiased, and by construction, it is inconsistent, because T cannot be assumed to be increasing. If that is the case, then $\hat{\overline{\phi}}$ is likewise inconsistent for $\overline{\phi}$. We are averaging n estimators, each of which has bias and variance that are $O(1/T)$. The variance of the mean is, therefore, $O(1/nT)$ which goes to zero, but the bias remains $O(1/T)$. It follows that the average of the n means is not converging to $\overline{\phi}$; it is converging to the average

11-99

of whatever these biased estimators are estimating. The problem vanishes with large T, but that is not relevant to the current context. However, in the Pesaran and Smith study, T was 29, which is large enough that these effects are probably moderate. For macroeconomic cross-country studies such as those based on the Penn World Tables, the data series might be yet longer than this.

One might consider aggregating the data to improve the results. Smith and Pesaran (1995) suggest an average based on country means. Averaging the observations over T in ~~(9-63)~~ 11-88 90 produces 91

$$\overline{y}_{i.} = \alpha_i + \beta_i \overline{x}_{i.} + \gamma_i \overline{y}_{-1,i} + \overline{\varepsilon}_{i.}. \qquad \text{11-89 } \cancel{(9\text{-}64)}$$

A linear regression using the n observations would be inconsistent for two reasons: First, $\overline{\varepsilon}_{i.}$ and $\overline{y}_{-1,i}$ must be correlated. Second, because of the parameter heterogeneity, it is not clear without further assumptions what the OLS slopes estimate under the false assumption that all coefficients are equal. But $\overline{y}_{i.}$ and $\overline{y}_{-1,i}$ differ by only the first and last observations; $\overline{y}_{-1,i} = \overline{y}_{i.} - (y_{iT} - y_{i0})/T = \overline{y}_{i.} - [\Delta_T(y)/T]$. Inserting this in ~~(9-64)~~ 11-89 produces

$$\begin{aligned}\overline{y}_{i.} &= \alpha_i + \beta_i \overline{x}_{i.} + \gamma_i \overline{y}_{i.} - \gamma_i[\Delta_T(y)/T] + \overline{\varepsilon}_{i.} \\ &= \frac{\alpha_i}{1-\gamma_i} + \frac{\beta_i}{1-\gamma_i}\overline{x}_{i.} - \frac{\gamma_i}{1-\gamma_i}[\Delta_T(y)/T] + \overline{\varepsilon}_{i.} \qquad \text{92 11-90 } \cancel{(9\text{-}65)} \\ &= \delta_i + \phi_i \overline{x}_{i.} + \tau_i[\Delta_T(y)/T] + \overline{\varepsilon}_{i.}.\end{aligned}$$

92 11-90

We still seek to estimate $\overline{\phi}$. The form in ~~(9-65)~~ does not solve the estimation problem, since the regression suggested using the group means is still heterogeneous. If it could be assumed that the individual long run coefficients differ randomly from the averages in the fashion of the random parameters model of the previous section, so $\delta_i = \overline{\delta} + u_{\delta,i}$ and likewise for the other parameters, then the model could be written

$$\begin{aligned}\overline{y}_{i.} &= \overline{\delta} + \overline{\phi}\overline{x}_{i.} + \overline{\tau}[\Delta_T(y)/T]_i + \overline{\varepsilon}_{i.} + \{u_{\delta,i} + u_{\phi,i}\overline{x}_i + u_{\tau,i}[\Delta_T(y)/T]_i\} \\ &= \overline{\delta} + \overline{\phi}\overline{x}_{i.} + \overline{\tau}[\Delta_T(y)/T]_i + \overline{\varepsilon}_i + w_i.\end{aligned}$$

At this point, the equation appears to be a heteroscedastic regression amenable to least squares estimation, but for one loose end. Consistency follows if the terms $[\Delta_T(y)/T]_i$ and $\overline{\varepsilon}_i$ are uncorrelated. Because the first is a rate of change and the second is in levels, this should generally be the case. Another interpretation that serves the same purpose is that the rates of change in $[\Delta_T(y)/T]_i$ should be uncorrelated with the levels in $\overline{x}_{i.}$, in which case, the regression can be partitioned, and simple linear regression of the country means of y_{it} on the country means of x_{it} and a constant produces consistent estimates of $\overline{\phi}$ and $\overline{\delta}$.

Alternatively, consider a time-series approach. We average the observation in ~~(9-63)~~ 90 11-88 across countries at each time period rather than across time within countries. In this case, we have

$$\overline{y}_{.t} = \overline{\alpha} + \frac{1}{n}\sum_{i=1}^{n}\beta_i x_{it} + \frac{1}{n}\sum_{i=1}^{n}\gamma_i y_{i,t-1} + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{it}.$$

Let $\overline{\gamma} = \frac{1}{n}\sum_{i=1}^{n}\gamma_i$ so that $\gamma_i = \overline{\gamma} + (\gamma_i - \overline{\gamma})$ and $\beta_i = \overline{\beta} + (\beta_i - \overline{\beta})$. Then,

$$\begin{aligned}\overline{y}_{.t} &= \overline{\alpha} + \overline{\beta}\overline{x}_{.t} + \overline{\gamma}\,\overline{y}_{-1,t} + [\overline{\varepsilon}_{.t} + (\beta_i - \overline{\beta})\overline{x}_{.t} + (\gamma_i - \overline{\gamma})\overline{y}_{-1,t}] \\ &= \overline{\alpha} + \overline{\beta}\overline{x}_{.t} + \overline{\gamma}\,\overline{y}_{-1,t} + \overline{\varepsilon}_{.t} + w_{.t}.\end{aligned}$$

11-100

Unfortunately, the regressor, $\overline{\gamma}\,\overline{y}_{-1,t}$ is surely correlated with $w_{.t}$, so neither OLS or GLS will provide a consistent estimator for this model. (One might consider an instrumental variable estimator, however, there is no natural instrument available in the model as constructed.) Another possibility is to pool the entire data set, possibly with random or fixed effects for the constant terms. Because pooling, even with country-specific constant terms, imposes homogeneity on the other parameters, the same problems we have just observed persist.

AU: Is "γ" a prime?

Finally, returning to (~~9-63~~ 11-88 90), one might treat it as a formal random parameters model,

$$
\begin{aligned}
y_{it} &= \alpha_i + \beta_i x_{it} + \gamma_i y_{i,t-1} + \varepsilon_{it},\\
\alpha_i &= \alpha + u_{\alpha,i},\\
\beta_i &= \beta + u_{\beta,i},\\
\gamma_i &= \gamma + u_{\gamma,i}.
\end{aligned}
\qquad (~~9\text{-}63'~~\ 11\text{-}88'\ 93')
$$

The assumptions needed to formulate the model in this fashion are those of the previous section. As Pesaran and Smith (1995) observe, this model can be estimated using the "Swamy (1971)" estimator, which is the matrix weighted average of the least squares estimators discussed in Section ~~9.8.1~~ 11.11.1. The estimator requires that T be large enough to fit each country regression by least squares. That has been the case for the received applications. Indeed, for the applications we have examined, both n and T are relatively large. If not, then one could still use the mixed models approach developed in ~~Section 9.8.2~~ Chapter 17. A compromise that appears to work well for panels with moderate sized n and T is the "mixed-fixed" model suggested in Hsiao (1986, 2003) and Weinhold (1999). The dynamic model in (~~9-63~~ 11-88 93) is formulated as a partial fixed effects model,

$$
\begin{aligned}
y_{it} &= \alpha_i d_{it} + \beta_i x_{it} + \gamma_i d_{it} y_{i,t-1} + \varepsilon_{it},\\
\beta_i &= \beta + u_{\beta,i},
\end{aligned}
$$

where d_{it} is a dummy variable that equals one for country i in every period and zero otherwise (i.e., the usual fixed effects approach). Note that d_{it} also appears with $y_{i,t-1}$. As stated, the model has "fixed effects," one random coefficient, and a total of $2n+1$ coefficients to estimate, in addition to the two variance components, σ_ε^2 and σ_u^2. The model could be estimated inefficiently by using ordinary least squares—the random coefficient induces heteroscedasticity (see Section ~~9.8.1~~ 11.11.1)—by using the Hildreth–Houck–Swamy approach, or with the mixed linear model approach developed in ~~Section 9.8.2~~ Chapter 17.

Example ~~9.19~~ 11.22 A Mixed Fixed Growth Model for Developing Countries

Weinhold (1996) and Nair–Reichert and Weinhold (2001) analyzed growth and development in a panel of 24 developing countries observed for 25 years, 1971–1995. The model they employed was a variant of the mixed-fixed model proposed by Hsiao (1986, 2003). In their specification,

$$
\begin{aligned}
GGDP_{i,t} = {} & \alpha_i d_{it} + \gamma_i d_{it} GGDP_{i,t-1}\\
& + \beta_{1i} GGDI_{i,t-1} + \beta_{2i} GFDI_{i,t-1} + \beta_{3i} GEXP_{i,t-1} + \beta_4 INFL_{i,t-1} + \varepsilon_{it},
\end{aligned}
$$

where

$GGDP$ = Growth rate of gross domestic product,
$GGDI$ = Growth rate of gross domestic investment,
$GFDI$ = Growth rate of foreign direct investment (inflows),
$GEXP$ = Growth rate of exports of goods and services,
$INFL$ = Inflation rate.

11-101

have at period $t = 3$

$$y_{i3} - y_{i2} = (\mathbf{x}_{i3} - \mathbf{x}_{i2})'\boldsymbol{\beta} + \gamma(y_{i2} - y_{i1}) + (\varepsilon_{i3} - \varepsilon_{i2}). \tag{9-72}$$

We could use y_{i1} as the needed variable, because it is not correlated $\varepsilon_{i3} - \varepsilon_{i2}$. Continuing in this fashion, we see that for $t = 3, 4, \ldots, T$, $y_{i,t-2}$ appears to satisfy our requirements. Alternatively, beginning from period $t = 4$, we can see that $z_{it} = (y_{i,t-2} - y_{i,t-3})$ once again satisfies our requirements. This is Anderson and Hsiao's (1981) result for instrumental variable estimation of the dynamic panel data model. It now becomes a question of which approach, levels $(y_{i,t-2}, t = 3, \ldots, T)$, or differences $(y_{i,t-2} - y_{i,t-3}, t = 4, \ldots, T)$ is a preferable approach. Kiviet (1995) obtains results that suggest that the estimator based on levels is more efficient.

This application has sketched the method of instrumental variables. There are numerous aspects yet to be considered, including a fuller development of the assumptions, the asymptotic distribution of the estimator, and what to use for an asymptotic covariance matrix to allow inference. We will return to the development of the method of instrumental variables in Chapter 12.

11.12 ~~9.10~~ SUMMARY AND CONCLUSIONS

This chapter ~~The preceding~~ has shown a few of the extensions of the classical model that can be obtained when panel data are available. In principle, any of the models we have examined before this chapter and all those we will consider later, including the multiple equation models, can be extended in the same way. The main advantage, as we noted at the outset, is that with panel data, one can formally model dynamic effects and the heterogeneity across groups that ~~is~~ are typical in microeconomic data.

We will find in Chapter 10 that to some extent this model of heterogeneity can be misleading. What might have appeared at one level to be differences in the variances of the disturbances across groups may well be due to heterogeneity of a different sort, associated with the coefficient vectors. We will consider this possibility in the next chapter. We will also examine some additional models for disturbance processes that arise naturally in a multiple equations context but are actually more general cases of some of the models we looked at earlier, such as the model of groupwise heteroscedasticity.

Key Terms and Concepts

- Adjustment equation
- Autocorrelation
- Balanced panel
- Between groups
- Cluster estimator
- Contiguity
- Contiguity matrix
- Contrasts
- Dynamic panel data model
- Equilibrium multiplier
- Error components model
- First difference
- Fixed effects
- Fixed panel
- Group means
- Group means estimator
- Hausman specification test
- Heterogeneity
- Hierarchical linear model
- Hierarchical model
- Individual effect
- Instrumental variable
- Instrumental variable estimator
- Lagrange multiplier test
- Least squares dummy variable estimator
- Long-run elasticity
- Long-run multiplier
- Longitudinal data sets
- Matrix weighted average
- Maximum simulated likelihood estimator
- Mean independence
- Measurement error
- Minimum distance estimator
- Mixed model
- Mundlak's approach

AU: The following terms were not bold KTs in chapter:
Hierachical linear model
Least squares dummy variable estimator
~~Maximum simulated likelihood estimator~~
~~Mixed model~~

See ms p 11-102

AU: The following terms were not bold KTs in chapter: ~~Nested random effects~~, Parameter heterogeneity, Pooled regression, ~~Random coefficients model~~, ~~Random parameters~~, Robust covariance matrix, ~~simulated log-likelihood~~, ~~Simulation based estimation~~, Small T asymptotics, Two-step estimation, Variable addition test

11-102

- Nested random effects
- Panel data
- Parameter heterogeneity
- Partial effects
- Pooled model
- Pooled regression
- Population averaged model
- Random coefficients model
- Random effects
- Random parameters
- Robust covariance matrix
- Rotating panel
- Simulated log-likelihood
- Simulation based estimation
- Small T asymptotics
- Spatial autocorrelation
- Spatial autoregression coefficient
- Spatial lags
- Specification test
- Strict exogeneity
- Time invariant
- Two-step estimation
- Unbalanced panel
- Variable addition test
- Within groups

Exercises

1. The following is a panel of data on investment (y) and profit (x) for $n = 3$ firms over $T = 10$ periods.

	$i=1$		$i=2$		$i=3$	
t	**y**	**x**	**y**	**x**	**y**	**x**
1	13.32	12.85	20.30	22.93	8.85	8.65
2	26.30	25.69	17.47	17.96	19.60	16.55
3	2.62	5.48	9.31	9.16	3.87	1.47
4	14.94	13.79	18.01	18.73	24.19	24.91
5	15.80	15.41	7.63	11.31	3.99	5.01
6	12.20	12.59	19.84	21.15	5.73	8.34
7	14.93	16.64	13.76	16.13	26.68	22.70
8	29.82	26.45	10.00	11.61	11.49	8.36
9	20.32	19.64	19.51	19.55	18.49	15.44
10	4.77	5.43	18.32	17.06	20.84	17.87

   a. Pool the data and compute the least squares regression coefficients of the model $y_{it} = \alpha + \beta x_{it} + \varepsilon_{it}$.
   b. Estimate the fixed effects model of ~~(9-12)~~, and then test the hypothesis that the constant term is the same for all three firms. 11-~~12~~ 13
   c. Estimate the random effects model of ~~(9-26)~~, and then carry out the Lagrange multiplier test of the hypothesis that the classical model without the common effect applies. 11-~~25~~ 29
   d. Carry out Hausman's specification test for the random versus the fixed effect model.

AU: Check style of Exercise titles. Elsewhere they were bold

2. Suppose that the fixed effects model is formulated with an overall constant term and $n - 1$ dummy variables (dropping, say, the last one). Investigate the effect that this supposition has on the set of dummy variable coefficients and on the least squares estimates of the slopes.
3. *Unbalanced design for random effects.* Suppose that the random effects model of Section 9.5 is to be estimated with a panel in which the groups have different numbers of observations. Let T_i be the number of observations in group i.
   a. Show that the pooled least squares estimator is unbiased and consistent despite this complication. 11-~~37~~ 40
   b. Show that the estimator in ~~(9-37)~~ based on the pooled least squares estimator of β (or, for that matter, *any* consistent estimator of β) is a consistent estimator of σ_ε^2.
4. What are the probability limits of $(1/n)$LM, where LM is defined in ~~(9-39)~~ under the null hypothesis that $\sigma_u^2 = 0$ and under the alternative that $\sigma_u^2 \neq 0$? 11-~~39~~ 42

11-103

5. *A two-way fixed effects model.* Suppose that the fixed effects model is modified to include a time-specific dummy variable as well as an individual-specific variable. Then $y_{it} = \alpha_i + \gamma_t + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$. At every observation, the individual- and time-specific dummy variables sum to 1, so there are some redundant coefficients. The discussion
11.4.4 in Section ~~9.4.4~~ shows that one way to remove the redundancy is to include an overall constant and drop one of the time specific *and* one of the time-dummy variables. The model is, thus,

$$y_{it} = \mu + (\alpha_i - \alpha_1) + (\gamma_t - \gamma_1) + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}.$$

(Note that the respective time- or individual-specific variable is zero when t or i equals one.) Ordinary least squares estimates of $\boldsymbol{\beta}$ are then obtained by regression of $y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{\bar{y}}$ on $\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.} - \bar{\mathbf{x}}_{.t} + \bar{\bar{\mathbf{x}}}$. Then $(\alpha_i - \alpha_1)$ and $(\gamma_t - \gamma_1)$ are estimated using the expressions in (9-23) while $m = \bar{\bar{y}} - \bar{\bar{\mathbf{x}}}'\mathbf{b}$. Using the following data, estimate the full set of coefficients for the least squares dummy variable model:

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	$t=9$	$t=10$
					$i=1$					
y	21.7	10.9	33.5	22.0	17.6	16.1	19.0	18.1	14.9	23.2
x_1	26.4	17.3	23.8	17.6	26.2	21.1	17.5	22.9	22.9	14.9
x_2	5.79	2.60	8.36	5.50	5.26	1.03	3.11	4.87	3.79	7.24
					$i=2$					
y	21.8	21.0	33.8	18.0	12.2	30.0	21.7	24.9	21.9	23.6
x_1	19.6	22.8	27.8	14.0	11.4	16.0	28.8	16.8	11.8	18.6
x_2	3.36	1.59	6.19	3.75	1.59	9.87	1.31	5.42	6.32	5.35
					$i=3$					
y	25.2	41.9	31.3	27.8	13.2	27.9	33.3	20.5	16.7	20.7
x_1	13.4	29.7	21.6	25.1	14.1	24.1	10.5	22.1	17.0	20.5
x_2	9.57	9.62	6.61	7.24	1.64	5.99	9.00	1.75	1.74	1.82
					$i=4$					
y	15.3	25.9	21.9	15.5	16.7	26.1	34.8	22.6	29.0	37.1
x_1	14.2	18.0	29.9	14.1	18.4	20.1	27.6	27.4	28.5	28.6
x_2	4.09	9.56	2.18	5.43	6.33	8.27	9.16	5.24	7.92	9.63

Test the hypotheses that (1) the "period" effects are all zero, (2) the "group" effects are all zero, and (3) both period and group effects are zero. Use an F test in each case.

6. *Two-way random effects model.* We modify the random effects model by the addition of a time-specific disturbance. Thus,

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} + u_i + v_t,$$

where

$$\begin{aligned}
E[\varepsilon_{it} \mid \mathbf{X}] &= E[u_i \mid \mathbf{X}] = E[v_t \mid \mathbf{X}] = 0, \\
E[\varepsilon_{it} u_j \mid \mathbf{X}] &= E[\varepsilon_{it} v_s \mid \mathbf{X}] = E[u_i v_t \mid \mathbf{X}] = 0 \quad \text{for all } i, j, t, s \\
\text{Var}[\varepsilon_{it} \mid \mathbf{X}] &= \sigma_\varepsilon^2, \quad \text{Cov}[\varepsilon_{it}, \varepsilon_{js} \mid \mathbf{X}] = 0 \quad \text{for all } i, j, t, s \\
\text{Var}[u_i \mid \mathbf{X}] &= \sigma_u^2, \quad \text{Cov}[u_i, u_j \mid \mathbf{X}] = 0 \quad \text{for all } i, j \\
\text{Var}[v_t \mid \mathbf{X}] &= \sigma_v^2, \quad \text{Cov}[v_t, v_s \mid \mathbf{X}] = 0 \quad \text{for all } t, s.
\end{aligned}$$

Write out the full disturbance covariance matrix for a data set with $n = 2$ and $T = 2$.

11-104

7. The model

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \beta + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}$$

satisfies the groupwise heteroscedastic regression model of Section ~~8.8.2.~~ All variables have zero means. The following sample second-moment matrix is obtained from a sample of 20 observations: 9.8.2

$$\begin{array}{c c} & \begin{array}{cccc} y_1 & y_2 & x_1 & x_2 \end{array} \\ \begin{array}{c} y_1 \\ y_2 \\ x_1 \\ x_2 \end{array} & \begin{bmatrix} 20 & 6 & 4 & 3 \\ 6 & 10 & 3 & 6 \\ 4 & 3 & 5 & 2 \\ 3 & 6 & 2 & 10 \end{bmatrix} \end{array}.$$

a. Compute the two separate OLS estimates of β, their sampling variances, the estimates of σ_1^2 and σ_2^2, and the R^2's in the two regressions.
b. Carry out the Lagrange multiplier test of the hypothesis that $\sigma_1^2 = \sigma_2^2$.
c. Compute the two-step FGLS estimate of β and an estimate of its sampling variance. Test the hypothesis that β equals 1.
d. Carry out the Wald test of equal disturbance variances.
e. Compute the maximum likelihood estimates of β, σ_1^2, and σ_2^2 by iterating the FGLS estimates to convergence.
f. Carry out a likelihood ratio test of equal disturbance variances.

8. Suppose that in the groupwise heteroscedasticity model of Section 8.8.2, $\mathbf{X}_i$ is the same for all i. What is the generalized least squares estimator of $\boldsymbol{\beta}$? How would you compute the estimator if it were necessary to estimate σ_i^2?
9. The following table presents a hypothetical panel of data:

	$i=1$		$i=2$		$i=3$	
t	**y**	**x**	**y**	**x**	**y**	**x**
1	30.27	24.31	38.71	28.35	37.03	21.16
2	35.59	28.47	29.74	27.38	43.82	26.76
3	17.90	23.74	11.29	12.74	37.12	22.21
4	44.90	25.44	26.17	21.08	24.34	19.02
5	37.58	20.80	5.85	14.02	26.15	18.64
6	23.15	10.55	29.01	20.43	26.01	18.97
7	30.53	18.40	30.38	28.13	29.64	21.35
8	39.90	25.40	36.03	21.78	30.25	21.34
9	20.44	13.57	37.90	25.65	25.41	15.86
10	36.85	25.60	33.90	11.66	26.04	13.28

a. Estimate the groupwise heteroscedastic model of Section ~~8.8.2~~. Include an estimate of the asymptotic variance of the slope estimator. Use a two-step procedure, basing the FGLS estimator at the second step on residuals from the pooled least squares regression. 9.8.2
b. Carry out the Wald and Lagrange multiplier tests of the hypothesis that the variances are all equal.

11-105

Applications

As usual, the following applications below require econometric software. The computations can be done with any modern software package, so no specific program is recommended.

1. The data in Appendix Table F~~9.3~~ 10.4 were used by Grunfeld (1958) and dozens of researchers since, including Zellner (1962, 1963) and Zellner and Huang (1962), to study different estimators for panel data and linear regression systems. The model is an investment equation [See Kleiber (2010) and Zeileis]

$$I_{it} = \beta_1 + \beta_2 F_{it} + \beta_3 C_{it} + \varepsilon_{it}, \; t = 1, \ldots, 20, \; i = 1, \ldots, 10,$$

where

$$I_{it} = \text{real gross investment for firm } i \text{ in year } t,$$
$$F_{it} = \text{real value of the firm—shares outstanding,}$$
$$C_{it} = \text{real value of the capital stock.}$$

For present purposes, this is a balanced panel data set.

a. Fit the pooled regression model.
b. Referring to the results in part a, is there evidence of within groups correlation? Compute the robust standard errors for your pooled OLS estimator and compare them to the conventional ones.
c. Compute the fixed effects estimator for these data, then, using an F test, test the hypothesis that the constants for the 10 firms are all the same.
d. Use a Lagrange multiplier statistic to test for the presence of common effects in the data.
e. Compute the one-way random effects estimator and report all estimation results. Explain the difference between this specification and the one in part c.
f. Use a Hausman test to determine whether a fixed or random effects specification is preferred for these data.

2. The data in Appendix Table F6.1 are an unbalanced panel on 25 U.S. airlines in the pre-deregulation days of the 1970s and 1980s. The group sizes range from 2 to 15. Data in the file are the following variables. (Variable names contained in the data file are constructed to indicate the variable contents.)

   Total cost,
   Expenditures on Capital, Labor, Fuel, Materials, Property, and Equipment,
   Price measures for the six inputs,
   Quantity measures for the six inputs,
   Output measured in revenue passenger miles, converted to an index number for the airline,
   Load factor = the average percentage capacity utilization of the airline's fleet,
   Stage = the average flight (stage) length in miles,
   Points = the number of points served by the airline,
   Year = the calendar year,
   T = Year—1969,
   TI = the number of observations for the airline, repeated for each year.

11-106
End 11

Use these data to build a cost model for airline service. Allow for cross-airline heterogeneity in the constants in the model. Use both random and fixed effects specifications, and use available statistical tests to determine which is the preferred model. An appropriate cost model to begin the analysis with would be

$$\ln cost_{it} = \alpha_i + \sum_{k=1}^{6} \beta_k \ln Price_{k,it} + \gamma \ln Output_{it} + \varepsilon_{it}.$$

It is necessary to impose linear homogeneity in the input prices on the cost function, which you would do by dividing five of the six prices and the total cost by the sixth price (choose any one), then using $\ln(cost/P_6)$ and $\ln(P_k/P_6)$ in the regression. You might also generalize the cost function by including a quadratic term in the log of output in the function. A translog model would include the unique squares and cross products of the input prices and products of log output with the logs of the prices. The data include three additional factors that may influence costs, stage length, load factor, and number of points served. Include them in your model, and use the appropriate test statistic to test whether they are, indeed, relevant to the determination of (log) total cost.