CHAPTER 12
Pooling Time-Series of Cross-Section Data

12.1 Introduction

In this chapter, we will consider pooling time-series of cross-sections. This may be a panel of households or firms or simply countries or states followed over time. Two well known examples of panel data in the U.S. are the Panel Study of Income Dynamics (PSID) and the National Longitudinal Survey (NLS). The PSID began in 1968 with 4802 families, including an over-sampling of poor households. Annual interviews were conducted and socioeconomic characteristics of each of the families and of roughly 31000 individuals who have been in these or derivative families were recorded. The list of variables collected is over 5000. The NLS, followed five distinct segments of the labor force. The original samples include 5020 older men, 5225 young men, 5083 mature women, 5159 young women and 12686 youths. There was an over-sampling of blacks, hispanics, poor whites and military in the youths survey. The list of variables collected runs into the thousands. An inventory of national studies using panel data is given at http://www.isr.umich.edu/src/psid/panelstudies.html. Pooling this data gives a richer source of variation which allows for more efficient estimation of the parameters. With additional, more informative data, one can get more reliable estimates and test more sophisticated behavioral models with less restrictive assumptions. Another advantage of panel data sets are their ability to control for individual heterogeneity. Not controlling for these unobserved individual specific effects leads to bias in the resulting estimates. Panel data sets are also better able to identify and estimate effects that are simply not detectable in pure cross-sections or pure time-series data. In particular, panel data sets are better able to study complex issues of dynamic behavior. For example, with a cross-section data set one can estimate the rate of unemployment at a particular point in time. Repeated cross-sections can show how this proportion changes over time. Only panel data sets can estimate what proportion of those who are unemployed in one period remain unemployed in another period. Some of the benefits and limitations of using panel data sets are listed in Hsiao (2003) and Baltagi (2008). Section 12.2 studies the error components model focusing on fixed effects, random effects and maximum likelihood estimation. Section 12.3 considers the question of prediction in a random effects model, while Section 12.4 illustrates the estimation methods using an empirical example. Section 12.5 considers testing the poolability assumption, the existence of random individual effects and the consistency of the random effects estimator using a Hausman test. Section 12.6 studies the dynamic panel data model and illustrates the methods used with an empirical example. Section 12.7 concludes with a short presentation of program evaluation and the difference-in-differences estimator.

12.2 The Error Components Model

The regression model is still the same, but it now has double subscripts

$$y_{it} = \alpha + X'_{it}\beta + u_{it} \tag{12.1}$$

B.H. Baltagi, *Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-20059-5_12,

where i denotes cross-sections and t denotes time-periods with $i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$. α is a scalar, β is $K \times 1$ and X_{it} is the it-th observation on K explanatory variables. The observations are usually stacked with i being the slower index, i.e., the T observations on the first household followed by the T observations on the second household, and so on, until we get to the N-th household. Under the error components specification, the disturbances take the form

$$u_{it} = \mu_i + \nu_{it} \tag{12.2}$$

where the μ_i's are cross-section specific components and ν_{it} are remainder effects. For example, μ_i may denote individual ability in an earnings equation, or managerial skill in a production function or simply a country specific effect. These effects are time-invariant.

In vector form, (12.1) can be written as

$$y = \alpha\iota_{NT} + X\beta + u = Z\delta + u \tag{12.3}$$

where y is $NT \times 1$, X is $NT \times K$, $Z = [\iota_{NT}, X]$, $\delta' = (\alpha', \beta')$, and ι_{NT} is a vector of ones of dimension NT. Also, (12.2) can be written as

$$u = Z_\mu\mu + \nu \tag{12.4}$$

where $u' = (u_{11}, \ldots, u_{1T}, u_{21}, \ldots, u_{2T}, \ldots, u_{N1}, \ldots, u_{NT})$ and $Z_\mu = I_N \otimes \iota_T$. I_N is an identity matrix of dimension N, ι_T is a vector of ones of dimension T, and $\otimes$ denotes Kronecker product defined in the Appendix to Chapter 7. Z_μ is a selector matrix of ones and zeros, or simply the matrix of individual dummies that one may include in the regression to estimate the μ_i's if they are assumed to be fixed parameters. $\mu' = (\mu_1, \ldots, \mu_N)$ and $\nu' = (\nu_{11}, \ldots, \nu_{1T}, \ldots, \nu_{N1}, \ldots, \nu_{NT})$. Note that $Z_\mu Z_\mu' = I_N \otimes J_T$ where J_T is a matrix of ones of dimension T, and $P = Z_\mu(Z_\mu' Z_\mu)^{-1}Z_\mu'$, the projection matrix on Z_μ, reduces to $P = I_N \otimes \bar{J}_T$ where $\bar{J}_T = J_T/T$. P is a matrix which averages the observation across time for each individual, and $Q = I_{NT} - P$ is a matrix which obtains the deviations from individual means. For example, Pu has a typical element $\bar{u}_{i.} = \sum_{t=1}^{T} u_{it}/T$ repeated T times for each individual and Qu has a typical element $(u_{it} - \bar{u}_{i.})$. P and Q are (i) symmetric *idempotent* matrices, i.e., $P' = P$ and $P^2 = P$. This means that the rank $(P) = \text{tr}(P) = N$ and rank $(Q) = \text{tr}(Q) = N(T-1)$. This uses the result that rank of an idempotent matrix is equal to its trace, see Graybill (1961, Theorem 1.63) and the Appendix to Chapter 7. Also, (ii) P and Q are *orthogonal*, i.e., $PQ = 0$ and (iii) they *sum to the identity matrix* $P + Q = I_{NT}$. In fact, any two of these properties imply the third, see Graybill (1961, Theorem 1.68).

12.2.1 The Fixed Effects Model

If the μ_i's are thought of as *fixed* parameters to be estimated, then equation (12.1) becomes

$$y_{it} = \alpha + X_{it}'\beta + \textstyle\sum_{i=1}^{N} \mu_i D_i + \nu_{it} \tag{12.5}$$

where D_i is a dummy variable for the i-th household. Not all the dummies are included so as not to fall in the dummy variable trap. One is usually dropped or equivalently, we can say that there is a restriction on the μ's given by $\sum_{i=1}^{N} \mu_i = 0$. The ν_{it}'s are the usual classical IID random variables with 0 mean and variance σ_ν^2. OLS on equation (12.5) is BLUE, but we have two problems, the first is the loss of degrees of freedom since in this case, we are estimating $N+K$ parameters. Also, with a lot of dummies we could be running into multicollinearity problems and a large $X'X$ matrix to invert. For example, if $N = 50$ states, $T = 10$ years and we have two explanatory variables, then with 500 observations we are estimating 52 parameters. Alternatively, we can think of this in an analysis of variance context and rearrange our observations, say, on y in an $(N \times T)$ matrix where rows denote firms and columns denote time periods.

		t				
		1	2	..	T	
	1	y_{11}	y_{12}	..	y_{1T}	$y_{1.}$
i	2	y_{21}	y_{22}	..	y_{2T}	$y_{2.}$
	:	:	:	..	:	:
	N	y_{N1}	y_{N2}	..	y_{NT}	$y_{N.}$

where $y_{i.} = \sum_{t=1}^{T} y_{it}$ and $\bar{y}_{i.} = y_{i.}/T$. For the simple regression with one regressor, the model given in (12.1) becomes

$$y_{it} = \alpha + \beta x_{it} + \mu_i + \nu_{it} \tag{12.6}$$

averaging over time gives

$$\bar{y}_{i.} = \alpha + \beta \bar{x}_{i.} + \mu_i + \bar{\nu}_{i.} \tag{12.7}$$

and averaging over all observations gives

$$\bar{y}_{..} = \alpha + \beta \bar{x}_{..} + \bar{\nu}_{..} \tag{12.8}$$

where $\bar{y}_{..} = \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it}/NT$. Equation (12.8) follows because the μ_i's sum to zero. Defining $\widetilde{y}_{it} = (y_{it} - \bar{y}_{i.})$ and $\widetilde{x}_{it}$ and $\widetilde{\nu}_{it}$ similarly, we get

$$y_{it} - \bar{y}_{i.} = \beta(x_{it} - \bar{x}_{i.}) + (\nu_{it} - \bar{\nu}_{i.})$$

or

$$\widetilde{y}_{it} = \beta \widetilde{x}_{it} + \widetilde{\nu}_{it} \tag{12.9}$$

Running OLS on equation (12.9) leads to the same estimator of β as that obtained from equation (12.5). This is called the least squares dummy variable estimator (LSDV) or $\widetilde{\beta}$ in our notation. It is also known as the *Within estimator* since $\sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{x}_{it}^2$ is the *within* sum of squares in an analysis of variance framework. One can then retrieve an estimate of α from equation (12.8) as $\widetilde{\alpha} = \bar{y}_{..} - \widetilde{\beta}\bar{x}_{..}$. Similarly, if we are interested in the μ_i's, those can also be retrieved from (12.7) and (12.8) as follows:

$$\widetilde{\mu}_i = (\bar{y}_{i.} - \bar{y}_{..}) - \widetilde{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \tag{12.10}$$

In matrix form, one can substitute the disturbances given by (12.4) into (12.3) to get

$$y = \alpha \iota_{NT} + X\beta + Z_\mu \mu + \nu = Z\delta + Z_\mu \mu + \nu \tag{12.11}$$

and then perform OLS on (12.11) to get estimates of α, β and μ. Note that Z is $NT \times (K+1)$ and Z_μ, the matrix of individual dummies is $NT \times N$. If N is large, (12.11) will include too many individual dummies, and the matrix to be inverted by OLS is large and of dimension $(N+K)$. In fact, since α and β are the parameters of interest, one can obtain the least squares dummy variables (LSDV) estimator from (12.11), by residualing out the Z_μ variables, i.e., by premultiplying the model by Q, the orthogonal projection of Z_μ, and performing OLS

$$Qy = QX\beta + Q\nu \tag{12.12}$$

This uses the fact that $QZ_\mu = Q\iota_{NT} = 0$, since $PZ_\mu = Z_\mu$. In other words, the Q matrix wipes out the individual effects. Recall, the FWL Theorem in Chapter 7. This is a regression of $\widetilde{y} = Qy$ with typical element $(y_{it} - \bar{y}_{i.})$ on $\widetilde{X} = QX$ with typical element $(X_{it,k} - \bar{X}_{i.,k})$ for the k-th regressor, $k = 1, 2, \ldots, K$. This involves the inversion of a $(K \times K)$ matrix rather than $(N+K) \times (N+K)$ as in (12.11). The resulting OLS estimator is

$$\widetilde{\beta} = (X'QX)^{-1}X'Qy \tag{12.13}$$

with $\text{var}(\widetilde{\beta}) = \sigma_\nu^2(X'QX)^{-1} = \sigma_\nu^2(\widetilde{X}'\widetilde{X})^{-1}$.

Note that this fixed effects (FE) estimator cannot estimate the effect of any time-invariant variable like sex, race, religion, schooling, or union participation. These time-invariant variables are wiped out by the Q transformation, the deviations from means transformation. Alternatively, one can see that these time-invariant variables are spanned by the individual dummies in (12.5) and therefore any regression package attempting (12.5) will fail, signaling perfect multicollinearity. If (12.5) is the true model, LSDV is BLUE as long as ν_{it} is the standard classical disturbance with mean 0 and variance covariance matrix $\sigma_\nu^2 I_{NT}$. Note that as $T \to \infty$, the FE estimator is consistent. However, if T is fixed and $N \to \infty$ as typical in short labor panels, then only the FE estimator of β is consistent, the FE estimators of the individual effects $(\alpha + \mu_i)$ are not consistent since the number of these parameters increase as N increases.

Testing for Fixed Effects: One could test the joint significance of these dummies, i.e., H_0; $\mu_1 = \mu_2 = .. = \mu_{N-1} = 0$, by performing an F-test. This is a simple Chow test given in (4.17) with the restricted residual sums of squares (RRSS) being that of OLS on the pooled model and the unrestricted residual sums of squares (URSS) being that of the LSDV regression. If N is large, one can perform the within transformation and use that residual sum of squares as the URSS. In this case

$$F_0 = \frac{(RRSS - URSS)/(N-1)}{URSS/(NT - N - K)} \overset{H_0}{\sim} F_{N-1,N(T-1)-K} \tag{12.14}$$

Computational Warning: One computational caution for those using the *Within* regression given by (12.12). The s^2 of this regression as obtained from a typical regression package divides the residual sums of squares by $NT - K$ since the intercept and the dummies are not included. The proper s^2, say s^{*2} from the LSDV regression in (12.5) would divide the same residual sums of squares by $N(T-1) - K$. Therefore, one has to adjust the variances obtained from the within regression (12.12) by multiplying the variance-covariance matrix by (s^{*2}/s^2) or simply by multiplying by $[NT - K]/[N(T-1) - K]$.

12.2.2 The Random Effects Model

There are too many parameters in the fixed effects model and the loss of degrees of freedom can be avoided if the μ_i's can be assumed random. In this case $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$, $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$

and the μ_i's are independent of the ν_{it}'s. In addition, the X_{it}'s are independent of the μ_i's and ν_{it}'s for all i and t. The random effects model is an appropriate specification if we are drawing N individuals randomly from a large population.

This specification implies a homoskedastic variance $\text{var}(u_{it}) = \sigma_\mu^2 + \sigma_\nu^2$ for all i and t, and an equi-correlated block-diagonal covariance matrix which exhibits serial correlation over time only between the disturbances of the same individual. In fact,

$$\begin{aligned} \text{cov}(u_{it}, u_{js}) &= \sigma_\mu^2 + \sigma_\nu^2 \quad \text{for} \quad i = j,\, t = s \\ &= \sigma_\mu^2 \qquad\quad \text{for} \quad i = j,\, t \neq s \end{aligned} \tag{12.15}$$

and zero otherwise. This also means that the correlation coefficient between u_{it} and u_{js} is

$$\begin{aligned} \rho &= \text{correl}(u_{it}, u_{js}) = 1 \quad \text{for} \quad i = j,\, t = s \\ &= \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\nu^2) \qquad \text{for} \quad i = j,\, t \neq s \end{aligned} \tag{12.16}$$

and zero otherwise. From (12.4), one can compute the variance-covariance matrix

$$\Omega = E(uu') = Z_\mu E(\mu\mu')Z_\mu' + E(\nu\nu') = \sigma_\mu^2(I_N \otimes J_T) + \sigma_\nu^2(I_N \otimes I_T) \tag{12.17}$$

In order to obtain the GLS estimator of the regression coefficients, we need Ω^{-1}. This is a huge matrix for typical panels and is of dimension $(NT \times NT)$. No brute force inversion should be attempted even if the researcher's application has a small N and T. For example, if we observe $N = 20$ firms over $T = 5$ time periods, Ω will be 100 by 100. We will follow a simple trick devised by Wansbeek and Kapteyn (1982) that allows the deviation of Ω^{-1} and $\Omega^{-1/2}$. Essentially, one replaces J_T by $T\bar{J}_T$, and I_T by $(E_T + \bar{J}_T)$ where E_T is by definition $(I_T - \bar{J}_T)$. In this case:

$$\Omega = T\sigma_\mu^2(I_N \otimes \bar{J}_T) + \sigma_\nu^2(I_N \otimes E_T) + \sigma_\nu^2(I_N \otimes \bar{J}_T)$$

collecting terms with the same matrices, we get

$$\Omega = (T\sigma_\mu^2 + \sigma_\nu^2)(I_N \otimes \bar{J}_T) + \sigma_\nu^2(I_N \otimes E_T) = \sigma_1^2 P + \sigma_\nu^2 Q \tag{12.18}$$

where $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$. (12.18) is the spectral decomposition representation of Ω, with σ_1^2 being the first unique characteristic root of Ω of multiplicity N and σ_ν^2 is the second unique characteristic root of Ω of multiplicity $N(T-1)$. It is easy to verify, using the properties of P and Q, that

$$\Omega^{-1} = \frac{1}{\sigma_1^2}P + \frac{1}{\sigma_\nu^2}Q \tag{12.19}$$

and

$$\Omega^{-1/2} = \frac{1}{\sigma_1}P + \frac{1}{\sigma_\nu}Q \tag{12.20}$$

In fact, $\Omega^r = (\sigma_1^2)^r P + (\sigma_\nu^2)^r Q$ where r is an arbitrary scalar. Now we can obtain GLS as a weighted least squares. Fuller and Battese (1974) suggested premultiplying the regression equation given in (12.3) by $\sigma_\nu\Omega^{-1/2} = Q + (\sigma_\nu/\sigma_1)P$ and performing OLS on the resulting transformed regression. In this case, $y^* = \sigma_\nu\Omega^{-1/2}y$ has a typical element $y_{it} - \theta\bar{y}_{i.}$ where $\theta = 1 - (\sigma_\nu/\sigma_1)$. This transformed regression inverts a matrix of dimension $(K+1)$ and can be easily implemented using any regression package.

The Best Quadratic Unbiased (BQU) estimators of the variance components arise naturally from the spectral decomposition of Ω. In fact, $Pu \sim (0, \sigma_1^2 P)$ and $Qu \sim (0, \sigma_\nu^2 Q)$ and

$$\widehat{\sigma}_1^2 = \frac{u'Pu}{\text{tr}(P)} = T\textstyle\sum_{i=1}^N \bar{u}_{i.}^2/N \tag{12.21}$$

and

$$\widehat{\sigma}_\nu^2 = \frac{u'Qu}{\text{tr}(Q)} = T\textstyle\sum_{i=1}^N \sum_{t=1}^T (u_{it} - \bar{u}_{i.})^2/N(T-1) \tag{12.22}$$

provide the BQU estimators of σ_1^2 and σ_ν^2, respectively, see Balestra (1973).

These are analysis of variance type estimators of the variance components and are MVU under normality of the disturbances, see Graybill (1961). The true disturbances are not known and therefore (12.21) and (12.22) are not feasible. Wallace and Hussain (1969) suggest substituting OLS residuals $\widehat{u}_{OLS}$ instead of the true u's. After all, the OLS estimates are still unbiased and consistent, but no longer efficient. Amemiya (1971) shows that these estimators of the variance components have a different asymptotic distribution from that knowing the true disturbances. He suggests using the LSDV residuals instead of the OLS residuals. In this case $\widetilde{u} = y - \widetilde{\alpha}\iota_{NT} - X\widetilde{\beta}$ where $\widetilde{\alpha} = \bar{y}_{..} - \bar{X}'_{..}\widetilde{\beta}$ and $\bar{X}'_{..}$ is a $1 \times K$ vector of averages of all regressors. Substituting these $\widetilde{u}$'s for u in (12.21) and (12.22) we get the Amemiya-type estimators of the variance components. The resulting estimates of the variance components have the same asymptotic distribution as that knowing the true disturbances.

Swamy and Arora (1972) suggest running two regressions to get estimates of the variance components from the corresponding mean square errors of these regressions. The first regression is the *Within* regression, given in (12.12), which yields the following s^2:

$$\widehat{\widehat{\sigma}}_\nu^2 = [y'Qy - y'QX(X'QX)^{-1}X'Qy]/[N(T-1) - K] \tag{12.23}$$

The second regression is the *Between* regression which runs the regression of averages across time, i.e.,

$$\bar{y}_{i.} = \alpha + \bar{X}'_{i.}\beta + \bar{u}_{i.} \qquad i = 1, \ldots, N \tag{12.24}$$

This is equivalent to premultiplying the model in (12.11) by P and running OLS. The only caution is that the latter regression has NT observations because it repeats the averages T times for each individual, while the cross-section regression in (12.24) is based on N observations. To remedy this, one can run the cross-section regression

$$y_{i.}/\sqrt{T} = \alpha(\sqrt{T}) + (X'_{i.}/\sqrt{T})\beta + u_{i.}/\sqrt{T} \tag{12.25}$$

where one can easily verify that $\text{var}(u_{i.}/\sqrt{T}) = \sigma_1^2$. This regression will yield an s^2 given by

$$\widehat{\widehat{\sigma}}_1^2 = (y'Py - y'PZ(Z'PZ)^{-1}Z'Py)/(N - K - 1) \tag{12.26}$$

Note that stacking the following two transformed regressions we just performed yields

$$\begin{pmatrix} Qy \\ Py \end{pmatrix} = \begin{pmatrix} QZ \\ PZ \end{pmatrix}\delta + \begin{pmatrix} Qu \\ Pu \end{pmatrix} \tag{12.27}$$

and the transformed error has mean 0 and variance-covariance matrix given by

$$\begin{pmatrix} \sigma_\nu^2 Q & 0 \\ 0 & \sigma_1^2 P \end{pmatrix}$$

Problem 6 asks the reader to verify that OLS on this system of $2NT$ observations yields OLS on the pooled model (12.3). Also, GLS on this system yields GLS on (12.3). Alternatively, one could get rid of the constant α by running the following stacked regressions:

$$\begin{pmatrix} Qy \\ (P - \bar{J}_{NT})y \end{pmatrix} = \begin{pmatrix} QX \\ (P - \bar{J}_{NT})X \end{pmatrix} \beta + \begin{pmatrix} Qu \\ (P - \bar{J}_{NT})u \end{pmatrix} \tag{12.28}$$

This follows from the fact the $Q\iota_{NT} = 0$ and $(P - \bar{J}_{NT})\iota_{NT} = 0$. The transformed error has zero mean and variance-covariance matrix

$$\begin{pmatrix} \sigma_\nu^2 Q & 0 \\ 0 & \sigma_1^2(P - \bar{J}_{NT}) \end{pmatrix} \tag{12.29}$$

OLS on this system, yields OLS on (12.3) and GLS on (12.28) yields GLS on (12.3). In fact,

$$\begin{aligned} \widehat{\beta}_{GLS} &= [(X'QX/\sigma_\nu^2) + X'(P - \bar{J}_{NT})X/\sigma_1^2]^{-1}[(X'Qy/\sigma_\nu^2) + (X'(P - \bar{J}_{NT})y/\sigma_1^2)] \\ &= [W_{XX} + \phi^2 B_{XX}]^{-1}[W_{Xy} + \phi^2 B_{Xy}] \end{aligned} \tag{12.30}$$

with $\mathrm{var}(\widehat{\beta}_{GLS}) = \sigma_\nu^2[W_{XX} + \phi^2 B_{XX}]^{-1}$. Note that $W_{XX} = X'QX$, $B_{XX} = X'(P - \bar{J}_{NT})X$ and $\phi^2 = \sigma_\nu^2/\sigma_1^2$. Also, the Within estimator of β is $\widetilde{\beta}_{Within} = W_{XX}^{-1}W_{Xy}$ and the Between estimator $\widehat{\beta}_{Between} = B_{XX}^{-1}B_{Xy}$. This shows that $\widehat{\beta}_{GLS}$ is a matrix weighted average of $\widetilde{\beta}_{Within}$ and $\widehat{\beta}_{Between}$ weighing each estimate by the inverse of its corresponding variance. In fact

$$\widehat{\beta}_{GLS} = W_1\widetilde{\beta}_{Within} + W_2\widehat{\beta}_{Between} \tag{12.31}$$

where $W_1 = [W_{XX} + \phi^2 B_{XX}]^{-1}W_{XX}$ and $W_2 = [W_{XX} + \phi^2 B_{XX}]^{-1}(\phi^2 B_{XX}) = I - W_1$. This was demonstrated by Maddala (1971). Note that (i) if $\sigma_\mu^2 = 0$, then $\phi^2 = 1$ and $\widehat{\beta}_{GLS}$ reduces to $\widehat{\beta}_{OLS}$. (ii) If $T \to \infty$, then $\phi^2 \to 0$ and $\widehat{\beta}_{GLS}$ tends to $\widetilde{\beta}_{Within}$. (iii) If $\phi^2 \to \infty$, then $\widehat{\beta}_{GLS}$ tends to $\widehat{\beta}_{Between}$. In other words, the Within estimator ignores the between variation, and the Between estimator ignores the within variation. The OLS estimator gives equal weight to the between and within variations. From (12.30), it is clear that $\mathrm{var}(\widetilde{\beta}_{Within})-$ $\mathrm{var}(\widehat{\beta}_{GLS})$ is a positive semi-definite matrix, since ϕ^2 is positive. However as $T \to \infty$ for any fixed N, $\phi^2 \to 0$ and both $\widehat{\beta}_{GLS}$ and $\widetilde{\beta}_{Within}$ have the same asymptotic variance.

Another estimator of the variance components was suggested by Nerlove (1971). His suggestion is to estimate $\widehat{\sigma}_\mu^2 = \sum_{i=1}^N(\widehat{\mu}_i - \bar{\widehat{\mu}})^2/(N-1)$ where $\widehat{\mu}_i$ are the dummy coefficients estimates from the LSDV regression. $\widehat{\sigma}_\nu^2$ is estimated from the within residual sums of squares divided by NT without correction for degrees of freedom.

Note that, except for Nerlove's (1971) method, one has to retrieve $\widehat{\sigma}_\mu^2$ as $(\widehat{\sigma}_1^2 - \widehat{\sigma}_\nu^2)/T$. In this case, there is no guarantee that the estimate of $\widehat{\sigma}_\mu^2$ would be non-negative. Searle (1971) has an extensive discussion of the problem of negative estimates of the variance components in the biometrics literature. One solution is to replace these negative estimates by zero. This in fact is the suggestion of the Monte Carlo study by Maddala and Mount (1973). This study finds that negative estimates occurred only when the true σ_μ^2 was small and close to zero. In these cases

OLS is still a viable estimator. Therefore, replacing negative $\widehat{\sigma}_{\mu}^{2}$ by zero is not a bad sin after all, and the problem is dismissed as not being serious.

Under the random effects model, GLS based on the true variance components is BLUE, and all the feasible GLS estimators considered are asymptotically efficient as either N or $T \to \infty$. Maddala and Mount (1973) compared OLS, Within, Between, feasible GLS methods, true GLS and MLE using their Monte Carlo study. They found little to choose among the various feasible GLS estimators in small samples and argued in favor of methods that were easier to compute.

Taylor (1980) derived exact finite sample results for the one-way error components model. He compared the Within estimator with the Swamy-Arora feasible GLS estimator. He found the following important results: (1) Feasible GLS is more efficient than FE for all but the fewest degrees of freedom. (2) The variance of feasible GLS is never more than 17% above the Cramér-Rao lower bound. (3) More efficient estimators of the variance components do not necessarily yield more efficient feasible GLS estimators. These finite sample results are confirmed by the Monte Carlo experiments carried out by Maddala and Mount (1973) and Baltagi (1981).

12.2.3 Maximum Likelihood Estimation

Under normality of the disturbances, one can write the log-likelihood function as

$$L(\alpha, \beta, \phi^2, \sigma_\nu^2) = \text{constant} - \frac{NT}{2}\log\sigma_\nu^2 + \frac{N}{2}\log\phi^2 - \frac{1}{2\sigma_\nu^2}u'\Sigma^{-1}u \tag{12.32}$$

where $\Omega = \sigma_\nu^2\Sigma$, $\phi^2 = \sigma_\nu^2/\sigma_1^2$ and $\Sigma = Q + \phi^{-2}P$ from (12.18). This uses the fact that $|\Omega| =$ product of its characteristic roots $= (\sigma_\nu^2)^{N(T-1)}(\sigma_1^2)^N = (\sigma_\nu^2)^{NT}(\phi^2)^{-N}$. Note that there is a one-to-one correspondence between ϕ^2 and σ_μ^2. In fact, $0 \leq \sigma_\mu^2 < \infty$ translates into $0 < \phi^2 \leq 1$. Brute force maximization of (12.32) leads to nonlinear first-order conditions, see Amemiya (1971). Instead, Breusch (1987) concentrates the likelihood with respect to α and σ_ν^2. In this case, $\widehat{\alpha}_{MLE} = \bar{y}_{..} - \bar{X}'_{..}\widehat{\beta}_{MLE}$ and $\widehat{\sigma}^2_{\nu,MLE} = \widehat{u}'\widehat{\Sigma}^{-1}\widehat{u}/NT$ where $\widehat{u}$ and $\widehat{\Sigma}$ are based on MLE's of β, ϕ^2 and α. Let $d = y - X\widehat{\beta}_{MLE}$ then $\widehat{\alpha}_{MLE} = \iota'_{NT}d/NT$ and $\widehat{u} = d - \iota_{NT}\widehat{\alpha} = d - \bar{J}_{NT}d$. This implies that $\widehat{\sigma}^2_{\nu,MLE}$ can be rewritten as

$$\widehat{\sigma}^2_{\nu,MLE} = d'[Q + \phi^2(P - \bar{J}_{NT})]d/NT \tag{12.33}$$

and the concentrated log-likelihood becomes

$$L_c(\beta, \phi^2) = \text{constant} - \frac{NT}{2}\log\{d'[Q + \phi^2(P - \bar{J}_{NT})]d\} + \frac{N}{2}\log\phi^2 \tag{12.34}$$

Maximizing (12.34), over ϕ^2 given β, yields

$$\widehat{\phi}^2 = \frac{d'Qd}{(T-1)d'(P - \bar{J}_{NT})d} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(d_{it} - \bar{d}_{i.})^2}{T(T-1)\sum_{i=1}^{N}(\bar{d}_{i.} - \bar{d}_{..})^2} \tag{12.35}$$

Maximizing (12.34) over β, given ϕ^2, yields

$$\widehat{\beta}_{MLE} = \left\{X'[Q + \phi^2(P - \bar{J}_{NT})]X\right\}^{-1} X'[Q + \phi^2(P - \bar{J}_{NT})]y \tag{12.36}$$

One can iterate between β and ϕ^2 until convergence. Breusch (1987) shows that provided $T > 1$, any i-th iteration β, call it β_i, gives $0 < \phi^2_{i+1} < \infty$ in the *(i+1)*th iteration. More importantly, Breusch (1987) shows that these ϕ_i^2's have a *remarkable property* of forming a monotonic sequence. In fact, starting from the Within estimator of β, for $\phi^2 = 0$, the next ϕ^2 is finite and positive and starts a monotonically increasing sequence of ϕ^2's. Similarly, starting from the Between estimator of β, for $(\phi^2 \to \infty)$ the next ϕ^2 is finite and positive and starts a monotonically decreasing sequence of ϕ^2's. Hence, to guard against the possibility of a local maximum, Breusch (1987) suggests starting with $\widetilde{\beta}_{Within}$ and $\widehat{\beta}_{Between}$ and iterating. If these two sequences converge to the same maximum, then this is the global maximum. If one starts with $\widehat{\beta}_{OLS}$ for $\phi^2 = 1$, and the next iteration obtains a larger ϕ^2, then we have a local maximum at the boundary $\phi^2 = 1$. Maddala (1971) finds that there are at most two maxima for the likelihood $L(\phi^2)$ for $0 < \phi^2 \leq 1$. Hence, we have to guard against one local maximum.

12.3 Prediction

Suppose we want to predict S periods ahead for the i-th individual. For the random effects model, the BLU estimator is GLS. Using the results in Chapter 9 on GLS, Goldberger's (1962) Best Linear Unbiased Predictor (BLUP) of $y_{i,T+S}$ is

$$\hat{y}_{i,T+S} = Z'_{i,T+S}\widehat{\delta}_{GLS} + w'\Omega^{-1}\widehat{u}_{GLS} \quad \text{for } S \geq 1 \tag{12.37}$$

where $\widehat{u}_{GLS} = y - Z\widehat{\delta}_{GLS}$ and $w = E(u_{i,T+S}u)$. Note that

$$u_{i,T+S} = \mu_i + \nu_{i,T+S} \tag{12.38}$$

and $w = \sigma_\mu^2(\ell_i \otimes \iota_T)$ where ℓ_i is the i-th column of I_N, i.e., ℓ_i is a vector that has 1 in the i-th position and zero elsewhere. In this case

$$w'\Omega^{-1} = \sigma_\mu^2(\ell_i' \otimes \iota_T')\left[\frac{1}{\sigma_1^2}P + \frac{1}{\sigma_\nu^2}Q\right] = \frac{\sigma_\mu^2}{\sigma_1^2}(\ell_i' \otimes \iota_T') \tag{12.39}$$

since $(\ell_i' \otimes \iota_T')P = (\ell_i' \otimes \iota_T')$ and $(\ell_i' \otimes \iota_T')Q = 0$. The typical element of $w'\Omega^{-1}\widehat{u}_{GLS}$ is $(T\sigma_\mu^2/\sigma_1^2)\widehat{\bar{u}}_{i.,GLS}$ where $\widehat{\bar{u}}_{i.,GLS} = \sum_{t=1}^{T}\widehat{u}_{it,GLS}/T$. Therefore, in (12.37), the BLUP for $y_{i,T+S}$ corrects the GLS prediction by a fraction of the mean of the GLS residuals corresponding to that i-th individual. This predictor was considered by Wansbeek and Kapteyn (1978) and Taub (1979).

12.4 Empirical Example

Baltagi and Griffin (1983) considered the following gasoline demand equation:

$$\log\frac{Gas}{Car} = \alpha + \beta_1\log\frac{Y}{N} + \beta_2\log\frac{P_{MG}}{P_{GDP}} + \beta_3\log\frac{Car}{N} + u \tag{12.40}$$

where Gas/Car is motor gasoline consumption per auto, Y/N is real income per capita, P_{MG}/P_{GDP} is real motor gasoline price and Car/N denotes the stock of cars per capita. This panel consists of annual observations across eighteen OECD countries, covering the period 1960–1978. The data for this example are provided on the Springer web site as GASOLINE.DAT. Table 12.1

gives the Stata output for the Within estimator using $xtreg, fe$. This is the regression described in (12.5) and computed as in (12.9). The Within estimator gives a low price elasticity for gasoline demand of -.322. The F-statistic for the significance of the country effects described in (12.14) yields an observed value of 83.96. This is distributed under the null as an $F(17, 321)$ and is statistically significant. This F-statistic is printed by Stata below the fixed effects output. In EViews, one invokes the test for redundant effects after running the fixed effects regression.

Table 12.1 Fixed Effects Estimator – Gasoline Demand Data

	Coef.	Std. Err.	T	$P > \|t\|$	[95% Conf. Interval]	
$\log(Y/N)$	0.6622498	0.073386	9.02	0.000	0.5178715	0.8066282
$\log(P_{MG}/P_{GDP})$	–0.3217025	0.0440992	–7.29	0.000	–0.4084626	–0.2349425
$\log(Car/N)$	–0.6404829	0.0296788	–21.58	0.000	–0.6988725	–0.5820933
Constant	2.40267	0.2253094	10.66	0.000	1.959401	2.84594
sigma_u	0.34841289					
sigma_e	0.09233034					
Rho	0.93438173	(fraction of variance due to u_i)				

Table 12.2 gives the Stata output for the Between estimator using $xtreg, be$.This is based on the regression given in (12.24). The Between estimator yields a high price elasticity of gasoline demand of -.964. These results were also verified using TSP.

Table 12.2 Between Estimator – Gasoline Demand Data

	Coef.	Std. Err.	T	$P > \|t\|$	[95% Conf. Interval]	
$\log(Y/N)$	0.9675763	0.1556662	6.22	0.000	0.6337055	1.301447
$\log(P_{MG}/P_{GDP})$	–0.9635503	0.1329214	–7.25	0.000	–1.248638	–0.6784622
$\log(Car/N)$	–0.795299	0.0824742	–9.64	0.000	–0.9721887	–0.6184094
Constant	2.54163	0.5267845	4.82	0.000	1.411789	3.67147

Table 12.3 gives the Stata output for the random effect model using $xtreg, re$. This is the Swamy and Arora (1972) estimator which yields a price elasticity of -.420. This is closer to the Within estimator than the Between estimator.

Table 12.3 Random Effects Estimator – Gasoline Demand Data

	Coef.	Std. Err.	T	$P > \|t\|$	[95% Conf. Interval]	
$\log(Y/N)$	0.5549858	0.0591282	9.39	0.000	0.4390967	0.6708749
$\log(P_{MG}/P_{GDP})$	–0.4203893	0.0399781	–10.52	0.000	–0.498745	–0.3420336
$\log(Car/N)$	–0.6068402	0.025515	–23.78	0.000	–0.6568487	–0.5568316
Constant	1.996699	0.184326	10.83	0.000	1.635427	2.357971
sigma_u	0.19554468					
sigma_e	0.09233034					
Rho	0.81769	(fraction of variance due to u_i)				

Table 12.4 Gasoline Demand Data. One-way Error Component Results

	β_1	β_2	β_3	ρ
OLS	0.890	−0.892	−0.763	0
	(0.036)*	(0.030)*	(0.019)*	
WALHUS	0.545	−0.447	−0.605	0.75
	(0.066)	(0.046)	(0.029)	
AMEMIYA	0.602	−0.366	−0.621	0.93
	(0.066)	(0.042)	(0.029)	
SWAR	0.555	−0.402	−0.607	0.82
	(0.059)	(0.042)	(0.026)	
IMLE	0.588	−0.378	−0.616	0.91
	(0.066)	(0.046)	(0.029)	

* These are biased standard errors when the true model has error component disturbances (see Moulton, 1986).
Source: Baltagi and Griffin (1983). Reproduced by permission of Elsevier Science Publishers B.V. (North-Holland).

Table 12.5 Gasoline Demand Data. Wallace and Hussain (1969) Estimator

Dependent Variable: GAS
Method: Panel EGLS (Cross-section random effects)

Sample: 1960 1978
Periods included: 19
Cross-sections included: 18
Total panel (balanced) observations: 342
Wallace and Hussain estimator of component variances

	Coefficient	Std. Error	t-Statistic	Prob.
C	1.938318	0.201817	9.604333	0.0000
$\log(Y/N)$	0.545202	0.065555	8.316682	0.0000
$\log(P_{MG}/P_{GDP})$	−0.447490	0.045763	−9.778438	0.0000
$\log(Car/N)$	−0.605086	0.028838	−20.98191	0.0000

Effects Specification

	S.D.	Rho
Cross-section random	0.196715	0.7508
Idiosyncratic random	0.113320	0.2492

Table 12.4 gives the parameter estimates for OLS and three feasible GLS estimates of the slope coefficients along with their standard errors, and the corresponding estimate of ρ defined in (12.16). These were obtained using EViews by invoking the random effects estimation on the individual effects and choosing the estimation method from the options menu. Breusch's (1987) iterative maximum likelihood was computed using Stata($xtreg, mle$) and TSP.

Table 12.5 gives the EViews output for the Wallace and Hussain (1969) random effects estimator, while Table 12.6 gives the EViews output for the Amemiya (1971) random effects estimator. Note that EViews calls the Amemiya estimator Wansbeek and Kapteyn (1989) since the latter paper generalizes this method to deal with unbalanced panels with missing observations, see Baltagi (2008) for details. Table 12.6 gives the Stata maximum likelihood output.

Table 12.6 Gasoline Demand Data. Wansbeek and Kapteyn (1989) Estimator

Dependent Variable: GAS
Method: Panel EGLS (Cross-section random effects)

Sample: 1960 1978
Periods included: 19
Cross-sections included: 18
Total panel (balanced) observations: 342
Wallace and Hussain estimator of component variances

	Coefficient	Std. Error	t-Statistic	Prob.
C	2.188322	0.216372	10.11372	0.0000
$\log(Y/N)$	0.601969	0.065876	9.137941	0.0000
$\log(P_{MG}/P_{GDP})$	−0.365500	0.041620	−8.781832	0.0000
$\log(Car/N)$	−0.620725	0.027356	−22.69053	0.0000

Effects Specification

	S.D.	Rho
Cross-section random	0.343826	0.9327
Idiosyncratic random	0.092330	0.0673

Table 12.7 Gasoline Demand Data. Random Effects Maximum Likelihood Estimator

. xtreg c y p car,mle

Random-effects ML regression	Number of obs	=	342
Group variable (i): coun	Number of groups	=	18
Random effects u_i ~ Gaussian	Obs per group: min	=	19
	avg	=	19.0
	max	=	19
	LR chi2(3)	=	609.75
Log likelihood = 282.47697	Prob > chi2	=	0.0000

c	Coef.	Std. Err.	z	$P>\|z\|$	[95% Conf.	Interval]
$\log(Y/N)$	.5881334	.0659581	8.92	0.000	.4588578	.717409
$\log(P_{MG}/P_{GDP})$	−.3780466	.0440663	−8.58	0.000	−.464415	−.2916782
$\log(Car/N)$	−.6163722	.0272054	−22.66	0.000	−.6696938	−.5630506
_cons	2.136168	.2156039	9.91	0.000	1.713593	2.558744
sigma_u	.2922939	.0545496			.2027512	.4213821
sigma_e	.0922537	.0036482			.0853734	.0996885
rho	.9094086	.0317608			.8303747	.9571561

Likelihood-ratio test of sigma_u = 0: chibar2(01)= 463.97 Prob >= chibar2 = 0.000

12.5 Testing in a Pooled Model

(1) The Chow-Test

Before pooling the data one may be concerned whether the data is poolable. This hypothesis is also known as the stability of the regression equation across firms or across time. It can be formulated in terms of an unrestricted model which involves a separate regression equation for each firm

$$y_i = Z_i\delta_i + u_i \quad \text{for } i = 1, 2, \ldots, N \tag{12.41}$$

where $y_i' = (y_{i1}, \ldots, y_{iT})$, $Z_i = [\iota_T, X_i]$ and X_i is $(T \times K)$. δ_i' is $1 \times (K+1)$ and u_i is $T \times 1$. The important thing to notice is that δ_i is different for every regional equation. We want to test the hypothesis H_0; δ_i=δ for all i, versus H_1; $\delta_i \neq \delta$ for some i. Under H_0 we can write the restricted model given in (12.41) as:

$$y = Z\delta + u \tag{12.42}$$

where $Z' = (Z_1', Z_2', \ldots, Z_N')$ and $u' = (u_1', u_2', \ldots, u_N')$. The unrestricted model can also be written as

$$y = \begin{pmatrix} Z_1 & 0 & \ldots & 0 \\ 0 & Z_2 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & Z_N \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_N \end{pmatrix} + u = Z^*\delta^* + u \tag{12.43}$$

where $\delta^{*\prime} = (\delta_1', \delta_2', \ldots, \delta_N')$ and $Z = Z^*I^*$ with $I^* = (\iota_N \otimes I_{K'})$, an $NK' \times K'$ matrix, with $K' = K + 1$. Hence the variables in Z are all linear combinations of the variables in Z^*. Under the assumption that $u \sim N(0, \sigma^2 I_{NT})$, the MVU estimator for δ in equation (12.42) is

$$\widehat{\delta}_{OLS} = \widehat{\delta}_{MLE} = (Z'Z)^{-1}Z'y \tag{12.44}$$

and therefore

$$y = Z\widehat{\delta}_{OLS} + e \tag{12.45}$$

implying that $e = (I_{NT} - Z(Z'Z)^{-1}Z')y = My = M(Z\delta + u) = Mu$ since MZ $= 0$. Similarly, under the alternative, the MVU for δ_i is given by

$$\widehat{\delta}_{i,OLS} = \widehat{\delta}_{i,MLE} = (Z_i'Z_i)^{-1}Z_i'y_i \tag{12.46}$$

and therefore

$$y_i = Z_i\widehat{\delta}_{i,OLS} + e_i \tag{12.47}$$

implying that $e_i = (I_T - Z_i(Z_i'Z_i)^{-1}Z_i')y_i = M_iy_i = M_i(Z_i\delta_i + u_i) = M_iu_i$ since $M_iZ_i = 0$, and this is true for $i = 1, 2, \ldots, N$. Also, let

$$M^* = I_{NT} - Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} = \begin{pmatrix} M_1 & 0 & \ldots & 0 \\ 0 & M_2 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & M_N \end{pmatrix}$$

One can easily deduce that $y = Z^*\hat{\delta}^* + e^*$ with $e^* = M^*y = M^*u$ and $\hat{\delta}^* = (Z^{*\prime}Z^*)^{-1}Z^{*\prime}y$. Note that both M and M^* are symmetric and idempotent with $MM^* = M^*$. This easily follows since

$$Z(Z'Z)^{-1}Z'Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} = Z(Z'Z)^{-1}I^{*\prime}Z^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} = Z(Z'Z)^{-1}Z'$$

This uses the fact that $Z = Z^*I^*$. Now, $e'e - e^{*\prime}e^* = u'(M - M^*)u$ and $e^{*\prime}e^* = u'M^*u$ are independent since $(M - M^*)M^* = 0$. Also, both quadratic forms when divided by σ^2 are distributed as χ^2 since $(M - M^*)$ and M^* are idempotent, see Judge et al. (1985). Dividing these quadratic forms by their respective degrees of freedom, and taking their ratio leads to the following test statistic:

$$\begin{aligned} F_{obs} &= \frac{(e'e - e^{*\prime}e^*)/(\text{tr}(M) - \text{tr}(M^*))}{e^{*\prime}e^*/\text{tr}(M^*)} \qquad (12.48) \\ &= \frac{(e'e - e_1'e_1 - e_2'e_2 - .. - e_N'e_N)/(N-1)K'}{(e_1'e_1 + e_2'e_2 + .. + e_N'e_N)/N(T-K')} \end{aligned}$$

Under H_0, F_{obs} is distributed as an $F((N-1)K', N(T-K'))$, see lemma 2.2 of Fisher (1970). This is exactly the Chow's (1960) test extended to the case of N linear regressions.

The URSS in this case is the sum of the N residual sum of squares obtained by applying OLS to (12.41), i.e., on each firm equation separately. The RRSS is simply the RSS from OLS performed on the pooled regression given by (12.42). In this case, there are $(N-1)K'$ restrictions and the URSS has $N(T-K')$ degrees of freedom. Similarly, one can test the stability of the regression across time. In this case, the degrees of freedom are $(T-1)K'$ and $N(T-K')$ respectively. Both tests target the whole set of regression coefficients including the constant. If the LSDV model is suspected to be the proper specification, then the intercepts are allowed to vary but the slopes remain the same. To test the stability of the slopes only, the same Chow-test can be utilized, however the RRSS is now that of the LSDV regression with firm (or time) dummies only. The number of restrictions becomes $(N-1)K$ for testing the stability of the slopes across firms and $(T-1)K$ for testing their stability across time.

The Chow-test however is proper under spherical disturbances, and if that hypothesis is not correct it will lead to improper inference. Baltagi (1981) showed that if the true specification of the disturbances is an error components structure then the Chow-test tend to reject poolability too often when in fact it is true. However, a generalization of the Chow-test which takes care of the general variance-covariance matrix is available in Zellner (1962). This is exactly the test of the null hypothesis H_0; $R\beta = r$ when Ω is that of the error components specification, see Chapter 9. Baltagi (1981) shows that this test performs well in Monte Carlo experiments. In this case, all we need to do is transform our model (under both the null and alternative hypotheses) such that the transformed disturbances have a variance of $\sigma^2 I_{NT}$, then apply the Chow-test on the transformed model. The later step is legitimate because the transformed disturbances have homoskedastic variances and the usual Chow-test is legitimate. Given $\Omega = \sigma^2\Sigma$, we premultiply the restricted model given in (12.42) by $\Sigma^{-1/2}$ and we call $\Sigma^{-1/2}y = \dot{y}$, $\Sigma^{-1/2}Z = \dot{Z}$ and $\Sigma^{-1/2}u = \dot{u}$. Hence

$$\dot{y} = \dot{Z}\delta + \dot{u} \qquad (12.49)$$

with $E(\dot{u}\dot{u}') = \Sigma^{-1/2}E(uu')\Sigma^{-1/2\prime} = \sigma^2 I_{NT}$. Similarly, we premultiply the unrestricted model given in (12.43) by $\Sigma^{-1/2}$ and we call $\Sigma^{-1/2}Z^* = \dot{Z}^*$. Therefore

$$\dot{y} = \dot{Z}^*\delta^* + \dot{u} \tag{12.50}$$

with $E(\dot{u}\dot{u}') = \sigma^2 I_{NT}$.

At this stage, we can test H_0; $\delta_i = \delta$ for every $i = 1, 2, \ldots, N$, simply by using the Chow-statistic, only now on the transformed models (12.49) and (12.50) since they satisfy $\dot{u} \sim N(0, \sigma^2 I_{NT})$. Note that $\dot{Z} = \dot{Z}^* I^*$ which is simply obtained from $Z = Z^* I^*$ by premultiplying by $\Sigma^{-1/2}$. Defining $\dot{M} = I_{NT} - \dot{Z}(\dot{Z}'\dot{Z})^{-1}\dot{Z}'$, and $\dot{M}^* = I_{NT} - \dot{Z}^*(\dot{Z}^{*\prime}\dot{Z}^*)^{-1}\dot{Z}^{*\prime}$, it is easy to show that $\dot{M}$ and $\dot{M}^*$ are both symmetric, idempotent and such that $\dot{M}\dot{M}^* = \dot{M}^*$. Once again the conditions for lemma 2.2 of Fisher (1970) are satisfied, and the test-statistic

$$\dot{F}_{obs} = \frac{(\dot{e}'\dot{e} - \dot{e}^{*\prime}\dot{e}^*)/(\text{tr}(\dot{M}) - \text{tr}(\dot{M}^*))}{\dot{e}^{*\prime}\dot{e}^*/\text{tr}(\dot{M}^*)} \sim F((N-1)K',\ N(T-K')) \tag{12.51}$$

where $\dot{e} = \dot{y} - \dot{Z}\widehat{\dot{\delta}}_{OLS}$ and $\widehat{\dot{\delta}}_{OLS} = (\dot{Z}'\dot{Z})^{-1}\dot{Z}'\dot{y}$ implying that $\dot{e} = \dot{M}\dot{y} = \dot{M}\dot{u}$. Similarly, $\dot{e}^* = \dot{y} - \dot{Z}^*\widehat{\dot{\delta}}^*_{OLS}$ and $\widehat{\dot{\delta}}^*_{OLS} = (\dot{Z}^{*\prime}\dot{Z}^*)^{-1}\dot{Z}^{*\prime}\dot{y}$ implying that $\dot{e}^* = \dot{M}^*\dot{y} = \dot{M}^*\dot{u}$. This is the Chow-test after premultiplying the model by $\Sigma^{-1/2}$ or simply applying the Fuller and Battese (1974) transformation. See Baltagi (2008) for details.

For the gasoline data in Baltagi and Griffin (1983), Chow's test for poolability across countries yields an observed F-statistic of 129.38 and is distributed as $F(68, 270)$ under H_0; $\delta_i = \delta$ for $i = 1, \ldots, N$. This tests the stability of four time-series regression coefficients across 18 countries. The unrestricted SSE is based upon 18 OLS time-series regressions, one for each country. For the stability of the slope coefficients only, H_0; $\beta_i = \beta$, an observed F-value of 27.33 is obtained which is distributed as $F(51, 270)$ under the null. Chow's test for poolability across time yields an F-value of 0.276 which is distributed as $F(72, 266)$ under H_0; $\delta_t = \delta$ for $t = 1, \ldots, T$. This tests the stability of four cross-section regression coefficients across 19 time periods. The unrestricted SSE is based upon 19 OLS cross-section regressions, one for each year. This does not reject poolability across time-periods. The test for poolability across countries, allowing for a one-way error components model yields an F-value of 21.64 which is distributed as $F(68, 270)$ under H_0; $\delta_i = \delta$ for $i = 1, \ldots, N$. The test for poolability across time yields an F-value of 1.66 which is distributed as $F(72, 266)$ under H_0; $\delta_t = \delta$ for $t = 1, \ldots, T$. This rejects H_0 at the 5% level.

(2) The Breusch-Pagan Test

Next, we look at a Lagrange Multiplier test developed by Breusch and Pagan (1980), which tests whether H_0; $\sigma^2_\mu = 0$. The test statistic is given by

$$LM = (NT/2(T-1)) \left[(\textstyle\sum_{i=1}^N e_{i.}^2 / \sum_{i=1}^N \sum_{t=1}^T e_{it}^2) - 1 \right]^2 \tag{12.52}$$

where e_{it} denotes the OLS residuals on the pooled model, $e_{i.}$ denote their sum over t, respectively. Under the null hypothesis H_0 this LM statistic is distributed as a χ^2_1. For the gasoline data in Baltagi and Griffin (1983), the Breusch and Pagan LM test yields an LM statistic of 1465.6. This is obtained using the Stata command *xtest0* after estimating the model with random effects. This is significant and rejects the null hypothesis. The corresponding likelihood ratio test assuming Normal disturbances is also reported by Stata maximum likelihood output for the random effects model. This yields an LR statistic of 463.97 which is asymptotically distributed as χ^2_1 under the null hypothesis H_0 and is also significant.

One problem with the Breusch-Pagan test is that it assumes that the alternative hypothesis is two-sided when we know that $\sigma_\mu^2 > 0$. A one-sided version of this test is given by Honda (1985):

$$HO = \sqrt{\frac{NT}{2(T-1)}}\left[\frac{e'(I_N \otimes J_T)e}{e'e} - 1\right] \overset{H_0}{\rightarrow} N(0,1) \tag{12.53}$$

where e denotes the vector of OLS residuals. Note that the square of this $N(0,1)$ statistic is the Breusch and Pagan (1980) LM test-statistic. Honda (1985) finds that this test statistic is *uniformly most powerful* and robust to non-normality. However, Moulton and Randolph (1989) showed that the asymptotic $N(0,1)$ approximation for this one-sided LM statistic can be poor even in large samples. They suggest an alternative Standardized Lagrange Multiplier (SLM) test whose asymptotic critical values are generally closer to the exact critical values than those of the LM test. This SLM test statistic centers and scales the one-sided LM statistic so that its mean is zero and its variance is one.

$$SLM = \frac{HO - E(HO)}{\sqrt{\text{var}(HO)}} = \frac{d - E(d)}{\sqrt{\text{var}(d)}} \tag{12.54}$$

where $d = e'De/e'e$ and $D = (I_N \otimes J_T)$. Using the results on moments of quadratic forms in regression residuals, see for e.g., Evans and King (1985), we get

$$E(d) = \text{tr}(D\bar{P}_Z)/p$$

and

$$\text{var}(d) = 2\{p\ \text{tr}(D\bar{P}_Z)^2 - [\text{tr}(D\bar{P}_Z)]^2\}/p^2(p+2) \tag{12.55}$$

where $p = n - (K+1)$ and $\bar{P}_Z = I_n - Z(Z'Z)^{-1}Z'$. Under the null hypothesis, SLM has an asymptotic $N(0,1)$ distribution.

(3) The Hausman-Test

A critical assumption in the error components regression model is that $E(u_{it}/X_{it}) = 0$. This is important given that the disturbances contain individual effects (the μ_i's) which are unobserved and may be correlated with the X_{it}'s. For example, in an earnings equation these μ_i's may denote unobservable ability of the individual and this may be correlated with the schooling variable included on the right hand side of this equation. In this case, $E(u_{it}/X_{it}) \neq 0$ and the GLS estimator $\widehat{\beta}_{GLS}$ becomes biased and inconsistent for β. However, the within transformation wipes out these μ_i's and leaves the Within estimator $\widetilde{\beta}_{Within}$ unbiased and consistent for β. Hausman (1978) suggests comparing $\widehat{\beta}_{GLS}$ and $\widetilde{\beta}_{Within}$, both of which are consistent under the null hypothesis H_0; $E(u_{it}/X_{it}) = 0$, but which will have different probability limits if H_0 is not true. In fact, $\widetilde{\beta}_{Within}$ is consistent whether H_0 is true or not, while $\widehat{\beta}_{GLS}$ is BLUE, consistent and asymptotically efficient under H_0, but is inconsistent when H_0 is false. A natural test statistic would be based on $\widehat{q} = \widehat{\beta}_{GLS} - \widetilde{\beta}_{Within}$. Under H_0, plim $\widehat{q} = 0$, and $\text{cov}(\widehat{q}, \widehat{\beta}_{GLS}) = 0$.

Using the fact that $\widehat{\beta}_{GLS} - \beta = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u$ and $\widetilde{\beta}_{Within} - \beta = (X'QX)^{-1}X'Qu$, one gets $E(\widehat{q}) = 0$ and

$$\begin{aligned}\text{cov}(\widehat{\beta}_{GLS}, \widehat{q}) &= \text{var}(\widehat{\beta}_{GLS}) - \text{cov}(\widehat{\beta}_{GLS}, \widetilde{\beta}_{Within}) \\ &= (X'\Omega^{-1}X)^{-1} - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(uu')QX(X'QX)^{-1} = 0\end{aligned}$$

Using the fact that $\widetilde{\beta}_{Within} = \widehat{\beta}_{GLS} - \widehat{q}$, one gets

$$\text{var}(\widetilde{\beta}_{Within}) = \text{var}(\widehat{\beta}_{GLS}) + \text{var}(\widehat{q}),$$

since $\text{cov}(\widehat{\beta}_{GLS}, \widehat{q}) = 0$. Therefore,

$$\text{var}(\widehat{q}) = \text{var}(\widetilde{\beta}_{Within}) - \text{var}(\widehat{\beta}_{GLS}) = \sigma_\nu^2 (X'QX)^{-1} - (X'\Omega^{-1}X)^{-1} \tag{12.56}$$

Hence, the Hausman test statistic is given by

$$m = \widehat{q}'[\text{var}(\widehat{q})]^{-1}\widehat{q} \tag{12.57}$$

and under H_0 is asymptotically distributed as χ_K^2, where K denotes the dimension of slope vector β. In order to make this test operational, Ω is replaced by a consistent estimator $\widehat{\Omega}$, and GLS by its corresponding FGLS. An alternative asymptotically equivalent test can be obtained from the augmented regression

$$y^* = X^*\beta + \widetilde{X}\gamma + w \tag{12.58}$$

where $y^* = \sigma_\nu \Omega^{-1/2} y$, $X^* = \sigma_\nu \Omega^{-1/2} X$ and $\widetilde{X} = QX$. Hausman's test is now equivalent to testing whether $\gamma = 0$. This is a standard Wald test for the omission of the variables $\widetilde{X}$ from (12.58).

This test was generalized by Arellano (1993) to make it robust to heteroskedasticity and autocorrelation of arbitrary forms. In fact, if either heteroskedasticity or serial correlation is present, the variances of the Within and GLS estimators are not valid and the corresponding Hausman test statistic is inappropriate. For the Baltagi and Griffin (1983) gasoline data, the Hausman test statistic based on the difference between the Within estimator and that of feasible GLS based on Swamy and Arora (1972) yields a χ_3^2 value of $m = 306.1$ which rejects the null hypothesis. This is obtained using the Stata command *hausman*.

12.6 Dynamic Panel Data Models

The dynamic error components regression is characterized by the presence of a lagged dependent variable among the regressors, i.e.,

$$y_{it} = \delta y_{i,t-1} + x_{it}'\beta + \mu_i + \nu_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{12.59}$$

where δ is a scalar, x_{it}' is $1 \times K$ and β is $K \times 1$. This model has been extensively studied by Anderson and Hsiao (1982). Since y_{it} is a function of μ_i, $y_{i,t-1}$ is also a function of μ_i. Therefore, $y_{i,t-1}$, a right hand regressor in (12.59), is correlated with the error term. This renders the OLS estimator biased and inconsistent even if the ν_{it}'s are not serially correlated. For the FE estimator, the within transformation wipes out the μ_i's, but $\tilde{y}_{i,t-1}$ will still be correlated with $\tilde{\nu}_{it}$ even if the ν_{it}'s are not serially correlated. In fact, the Within estimator will be biased of $O(1/T)$ and its consistency will depend upon T being large, see Nickell (1981). An alternative transformation that wipes out the individual effects, yet does not create the above problem is the first difference (FD) transformation. In fact, Anderson and Hsiao (1982) suggested first differencing the model to get rid of the μ_i's and then using $\Delta y_{i,t-2} = (y_{i,t-2} - y_{i,t-3})$ or simply $y_{i,t-2}$ as an instrument for $\Delta y_{i,t-1} = (y_{i,t-1} - y_{i,t-2})$. These instruments will not be correlated with $\Delta\nu_{it} = \nu_{i,t} - \nu_{i,t-1}$, as long as the ν_{it}'s themselves are not serially correlated. This instrumental variable (IV) estimation method leads to consistent but not necessarily efficient estimates of the parameters in the model. This is because it does not make use of all the available moment conditions, see Ahn and Schmidt (1995), and it does not take into account the differenced structure on the residual disturbances ($\Delta\nu_{it}$). Arellano (1989) finds that for simple dynamic error components models the estimator that uses differences $\Delta y_{i,t-2}$ rather than levels $y_{i,t-2}$ for instruments has a singularity point and very large variances over a significant range of parameter values. In contrast, the estimator that uses instruments in levels, i.e., $y_{i,t-2}$, has no singularities and much smaller variances and is therefore recommended. Additional instruments can be obtained in a dynamic panel data model if one utilizes the orthogonality conditions that exist between lagged values of y_{it} and the disturbances ν_{it}.

Let us illustrate this with the simple autoregressive model with no regressors:

$$y_{it} = \delta y_{i,t-1} + u_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T \tag{12.60}$$

where $u_{it} = \mu_i + \nu_{it}$ with $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ and $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$, independent of each other and among themselves. In order to get a consistent estimate of δ as $N \to \infty$ with T fixed, we first difference (12.60) to eliminate the individual effects

$$y_{it} - y_{i,t-1} = \delta(y_{i,t-1} - y_{i,t-2}) + (\nu_{it} - \nu_{i,t-1}) \tag{12.61}$$

and note that $(\nu_{it} - \nu_{i,t-1})$ is MA(1) with unit root. For the first period we observe this relationship, i.e., $t = 3$, we have

$$y_{i3} - y_{i2} = \delta(y_{i2} - y_{i1}) + (\nu_{i3} - \nu_{i2})$$

In this case, y_{i1} is a valid instrument, since it is highly correlated with $(y_{i2} - y_{i1})$ and not correlated with $(\nu_{i3} - \nu_{i2})$ as long as the ν_{it} are not serially correlated. But note what happens for $t = 4$, the second period we observe (12.61):

$$y_{i4} - y_{i3} = \delta(y_{i3} - y_{i2}) + (\nu_{i4} - \nu_{i3})$$

In this case, y_{i2} as well as y_{i1} are valid instruments for $(y_{i3} - y_{i2})$, since both y_{i2} and y_{i1} are not correlated with $(\nu_{i4} - \nu_{i3})$. One can continue in this fashion, adding an extra valid instrument with each forward period, so that for period T, the set of valid instruments becomes $(y_{i1}, y_{i2}, \ldots, y_{i,T-2})$.

This instrumental variable procedure still does not account for the differenced error term in (12.61). In fact,

$$E(\Delta\nu_i \, \Delta\nu_i') = \sigma_\nu^2 G \tag{12.62}$$

where $\Delta\nu_i' = (\nu_{i3} - \nu_{i2}, \ldots, \nu_{iT} - \nu_{i,T-1})$ and

$$G = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix}$$

is $(T-2)\times(T-2)$, since $\Delta\nu_i$ is MA(1) with unit root. Define

$$W_i = \begin{bmatrix} [y_{i1}] & & & 0 \\ & [y_{i1}, y_{i2}] & & \\ & & \ddots & \\ 0 & & & [y_{i1}, \ldots, y_{i,T-2}] \end{bmatrix} \tag{12.63}$$

Then, the matrix of instruments is $W = [W_1', \ldots, W_N']'$ and the moment equations described above are given by $E(W_i'\Delta\nu_i) = 0$. Premultiplying the differenced equation (12.61) in vector form by W', one gets

$$W'\Delta y = W'(\Delta y_{-1})\delta + W'\Delta\nu \tag{12.64}$$

Performing GLS on (12.64) one gets the Arellano and Bond (1991) preliminary one-step consistent estimator

$$\begin{aligned}\widehat{\delta}_1 \;=\; & [(\Delta y_{-1})'W(W'(I_N\otimes G)W)^{-1}W'(\Delta y_{-1})]^{-1} \\ & \times[(\Delta y_{-1})'W(W'(I_N\otimes G)W)^{-1}W'(\Delta y)]\end{aligned} \tag{12.65}$$

The optimal generalized method of moments (GMM) estimator of δ_1 à la Hansen (1982) for $N\to\infty$ and T fixed using only the above moment restrictions yields the same expression as in (12.65) except that

$$W'(I_N\otimes G)W = \sum_{i=1}^{N} W_i'GW_i$$

is replaced by

$$V_N = \sum_{i=1}^{N} W_i'(\Delta\nu_i)(\Delta\nu_i)'W_i$$

This GMM estimator requires no knowledge concerning the initial conditions or the distributions of ν_i and μ_i. To operationalize this estimator, $\Delta\nu$ is replaced by differenced residuals obtained from the preliminary consistent estimator $\widehat{\delta}_1$. The resulting estimator is the two-step Arellano and Bond (1991) GMM estimator:

$$\widehat{\delta}_2 = [(\Delta y_{-1})'W\widehat{V}_N^{-1}W'(\Delta y_{-1})]^{-1}[(\Delta y_{-1})'W\widehat{V}_N^{-1}W'(\Delta y)] \tag{12.66}$$

A consistent estimate of the asymptotic var$(\widehat{\delta}_2)$ is given by the first term in (12.66),

$$\widehat{\mathrm{var}}(\widehat{\delta}_2) = [(\Delta y_{-1})'W\widehat{V}_N^{-1}W'(\Delta y_{-1})]^{-1} \tag{12.67}$$

Note that $\widehat{\delta}_1$ and $\widehat{\delta}_2$ are asymptotically equivalent if the ν_{it} are IID$(0,\sigma_\nu^2)$.

If there are additional *strictly exogenous* regressors x_{it} as in (12.59) with $E(x_{it}\nu_{is}) = 0$ for all $t, s = 1, 2, \ldots, T$, but where all the x_{it} are correlated with μ_i, then all the x_{it} are valid instruments for the first differenced equation of (12.59). Therefore, $[x_{i1}', x_{i2}', \ldots, x_{iT}']$ should be added to each diagonal element of W_i in (12.63). In this case, (12.64) becomes

$$W'\Delta y = W'(\Delta y_{-1})\delta + W'(\Delta X)\beta + W'\Delta\nu$$

where ΔX is the stacked $N(T-2) \times K$ matrix of observations on Δx_{it}. One- and two-step estimators of (δ, β')can be obtained from

$$\begin{pmatrix} \widehat{\delta} \\ \widehat{\beta} \end{pmatrix} = ([\Delta y_{-1}, \Delta X]' W \widehat{V}_N^{-1} W' [\Delta y_{-1}, \Delta X])^{-1} ([\Delta y_{-1}, \Delta X]' W \widehat{V}_N^{-1} W' \Delta y) \tag{12.68}$$

as in (12.65) and (12.66).

Arellano and Bond (1991) suggest Sargan's (1958) test of over-identifying restrictions given by

$$m = (\Delta\widehat{\nu})' W \left[\sum_{i=1}^{N} W_i' (\Delta\widehat{\nu}_i)(\Delta\widehat{\nu}_i)' W_i \right]^{-1} W'(\Delta\widehat{\nu}) \sim \chi^2_{p-K-1}$$

where p refers to the number of columns of W and $\Delta\widehat{\nu}$ denote the residuals from a two-step estimation given in (12.68).

To summarize, dynamic panel data estimation of equation (12.59) with individual fixed effects suffers from the Nickell (1981) bias. This disappears only if T tends to infinity. Alternatively, a GMM estimator was suggested by Arellano and Bond (1991) which basically differences the model to get rid of the individual specific effects and along with it any time invariant regressor. This also gets rid of any endogeneity that may be due to the correlation of these individual effects and the right hand side regressors. The moment conditions utilize the orthogonality conditions between the differenced errors and lagged values of the dependent variable. This assumes that the original disturbances are serially uncorrelated. In fact, two diagnostics are computed using the Arellano and Bond GMM procedure to test for first order and second order serial correlation in the disturbances. One should reject the null of the absence of first order serial correlation and not reject the absence of second order serial correlation. A special feature of dynamic panel data GMM estimation is that the number of moment conditions increase with T. Therefore, a Sargan test is performed to test the over-identification restrictions. There is convincing evidence that too many moment conditions introduce bias while increasing efficiency. It is even suggested that a subset of these moment conditions be used to take advantage of the trade-off between the reduction in bias and the loss in efficiency, see Baltagi (2008) for details.

Arellano and Bond (1991) apply their GMM estimation and testing methods to a model of employment using a panel of 140 quoted UK companies for the period 1979–84. This is the benchmark data set used in Stata to obtain the one-step and two-step estimators described in (12.65) and (12.66) as well as the Sargan test for over-identification using the command (xtabond,twostep).The reader is asked to replicate their results in problem 22.

12.6.1 Empirical Illustration

Baltagi, Griffin and Xiong (2000) estimate a dynamic demand model for cigarettes based on panel data from 46 American states over 30 years 1963–1992. The estimated equation is

$$\ln C_{it} = \alpha + \beta_1 \ln C_{i,t-1} + \beta_2 \ln P_{i,t} + \beta_3 \ln Y_{it} + \beta_4 \ln Pn_{it} + u_{it} \tag{12.69}$$

where the subscript i denotes the ith state $(i = 1, \ldots, 46)$, and the subscript t denotes the tth year $(t = 1, \ldots, 30)$. C_{it} is real per capita sales of cigarettes by persons of smoking age (14 years and older). This is measured in packs of cigarettes per head. P_{it} is the average retail price of a pack of cigarettes measured in real terms. Y_{it} is real per capita disposable income. Pn_{it} denotes the minimum real price of cigarettes in any neighboring state. This last variable is a proxy for the casual smuggling effect across state borders. It acts as a substitute price attracting consumers from high-tax states like Massachusetts to cross over to New Hampshire where the tax is low. The disturbance term is specified as a two-way error component model:

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \quad i = 1, \ldots, 46 \quad t = 1, \ldots, 30 \tag{12.70}$$

where μ_i denotes a state-specific effect, and λ_t denotes a year-specific effect. The time-period effects (the λ_t) are assumed fixed parameters to be estimated as coefficients of time dummies for each year in the sample. This can be justified given the numerous policy interventions as well as health warnings and Surgeon General's reports. For example:

(1) the imposition of warning labels by the Federal Trade Commission effective January 1965;

(2) the application of the Fairness Doctrine Act to cigarette advertising in June 1967, which subsidized antismoking messages from 1968 to 1970;

(3) the Congressional ban on broadcast advertising of cigarettes effective January 1971.

The μ_i are state-specific effects which can represent any state-specific characteristic including the following:

(1) States with Indian reservations like Montana, New Mexico and Arizona are among the biggest losers in tax revenues from non-Indians purchasing tax-exempt cigarettes from the reservations.

(2) Florida, Texas, Washington and Georgia are among the biggest losers of revenues due to the purchasing of cigarettes from tax-exempt military bases in these states.

(3) Utah, which has a high percentage of Mormon population (a religion which forbids smoking), has a per capita sales of cigarettes in 1988 of 55 packs, a little less than half the national average of 113 packs.

(4) Nevada, which is a highly touristic state, has a per capita sales of cigarettes of 142 packs in 1988, 29 more packs than the national average.

These state-specific effects may be assumed fixed, in which case one includes state dummy variables in equation (12.69). The resulting estimator is the Within estimator reported in Table 12.8. Comparing these estimates with OLS without state or time dummies, one can see that the coefficient of lagged consumption drops from 0.97 to 0.83 and the price elasticity goes up in absolute value from −0.09 to −0.30. The income elasticity switches sign from negative to positive going from −0.03 to 0.10.

The OLS and Within estimators do not take into account the endogeneity of the lagged dependent variable, and therefore 2SLS and Within-2SLS are performed. The instruments used are one lag on price, neighboring price and income. These give lower estimates of lagged consumption and higher own price elasticities in absolute value. The Arellano and Bond (1991) two-step estimator yields an estimate of lagged consumption of 0.70 and a price elasticity of −0.40, both of which are significant. Sargan's test for over-identification yields an observed value of 32.3. This is asymptotically distributed as χ^2_{27} and is not significant. This was obtained using the Stata command ($xtabond2, twostep$) with the collapse option to reduce the number of moment conditions used for estimation.

Table 12.8 Dynamic Demand for Cigarettes: 1963–92*

	$lnC_{i,t-1}$	lnP_{it}	lnY_{it}	$lnPn_{it}$
OLS	0.97	−0.090	−0.03	0.024
	(157.7)	(6.2)	(5.1)	(1.8)
Within	0.83	−0.299	0.10	0.034
	(66.3)	(12.7)	(4.2)	(1.2)
2SLS	0.85	−0.205	−0.02	0.052
	(25.3)	(5.8)	(2.2)	(3.1)
Within-2SLS	0.60	−0.496	0.19	−0.016
	(17.0)	(13.0)	(6.4)	(0.5)
Arellano and Bond (two-step)	0.70	−0.396	0.13	−0.003
	(10.2)	(6.0)	(3.5)	(0.1)

* Numbers in parentheses are t-statistics.
Source: Some of the results in this Table are reported in Baltagi, Griffin and Xiong (2000).

12.7 Program Evaluation and Difference-in-Differences Estimator

Suppose we want to study the effect of job training programs on earnings. An ideal experiment would assign individuals randomly, by a flip of a coin, to training and non-training camps, and then compare their earnings, holding other factors constant. This is a necessary experiment before the approval of any drug. Patients are randomly assigned to receive the drug or a placebo and the drug is approved or disapproved depending on the difference in the outcome between these two groups. In this case, the FDA is concerned with the drug's safety and its effectiveness. However, we run into problems in setting this experiment. How can we hold other factors constant? Even twins which have been used in economic studies are not identical and may have different life experiences.

The individual's prior work experience will affect one's chances in getting a job after training. But as long as the individuals are randomly assigned, the distribution of work experience is the same in the treatment and control group, i.e., participation in the job training is independent of prior work experience. In this case, omitting previous work experience from the analysis will not cause omitted variable bias in the estimator of the effect of the training program on future employment. Stock and Watson (2003) discuss threats to the internal and external validity of such experiments. The former include: (i) failure to randomize, or (ii) to follow the treatment protocol. These failures can cause bias in estimating the effect of the treatment. The first can happen when individuals are assigned non-randomly to the treatment and non-treatment groups. The second can happen, for example, when some people in the training program do not show up for all training sessions; or when some people who are not supposed to be in the training program are allowed to attend some of these training sessions. Attrition caused by people dropping out of the experiment in either group can cause bias especially if the cause of attrition is related to their acquiring or not acquiring training. In addition, small samples, usually associated with expensive experiments, can affect the precision of the estimates. There can also be experimental effects, brought about by people trying harder simply because the worker being trained feels noticed or because the trainer has a stake in the success of the program. Stock and Watson (2003, p. 380) argue that "threats to external validity compromise the ability to generalize the results of the experiment to other populations and settings. Two such threats are when the experimental sample is not representative of the population of interest and when the treatment being studied is not representative of the treatment that would be implemented more broadly."

They also warn about "general equilibrium effects" where, for example, turning a small, temporary experimental program into a widespread, permanent program might change the economic environment sufficiently that the results of the experiment cannot be generalized. For example, it could displace employer-provided training, thereby reducing the net benefits of the program.

12.7.1 The Difference-in-Differences Estimator

With panel data, observations on the same subjects before and after the training program allow us to estimate the effect of this program on earnings. In simple regression form, assuming the assignment to the training program is random, one regresses the change in earnings before and after training is completed on a dummy variable which takes the value 1 if the individual received training and zero if they did not. This regression computes the average change in earnings for the treatment group before and after the training program and subtracts that from the average change in earnings for the control group. One can include additional regressors which measure the individual characteristics prior to training. Examples are gender, race, education and age of the individual.

Card (1990) used a quasi-experiment to see whether immigration reduces wages. Taking advantage of the "Mariel boatlift" where a large number of Cuban immigrants entered Miami. Card (1990) used the difference-in-differences estimator, comparing the change in wages of low-skilled workers in Miami to the change in wages of similar workers in other comparable U.S. cities over the same period. Card concluded that the influx of Cuban immigrants had a negligible effect on wages of less-skilled workers.

Problems

1. *Fixed Effects and the Within Transformation.*

   (a) Premultiply (12.11) by Q and verify that the transformed equation reduces to (12.12). Show that the new disturbances $Q\nu$ have zero mean and variance-covariance matrix $\sigma_\nu^2 Q$.
   Hint: $QZ_\mu = 0$.

   (b) Show that the GLS estimator is the same as the OLS estimator on this transformed regression equation. **Hint:** Use one of the necessary and sufficient conditions for GLS to be equivalent to OLS given in Chapter 9.

   (c) Using the Frisch-Waugh-Lovell Theorem given in Chapter 7, show that the estimator derived in part (b) is the Within estimator and is given by $\widetilde{\beta} = (X'QX)^{-1}X'Qy$.

2. *Variance-Covariance Matrix of Random Effects.*

   (a) Show that Ω given in (12.17) can be written as (12.18).

   (b) Show that P and Q are symmetric, idempotent, orthogonal and sum to the identity matrix.

   (c) For Ω^{-1} given by (12.19), verify that $\Omega\Omega^{-1} = \Omega^{-1}\Omega = I_{NT}$.

   (d) For $\Omega^{-1/2}$ given by (12.20), verify that $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$.

3. *Fuller and Battese (1974) Transformation.* Premultiply y by $\sigma_\nu \Omega^{-1/2}$ where $\Omega^{-1/2}$ is defined in (12.20) and show that the resulting y^* has a typical element $y^*_{it} = y_{it} - \theta \bar{y}_{i.}$, where the $\theta = 1 - \sigma_\nu/\sigma_1$ and $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$.

4. *Unbiased Estimates of the Variance-Components.* Using (12.21) and (12.22), show that $E(\widehat{\sigma}_1^2) = \sigma_1^2$ and $E(\widehat{\sigma}_\nu^2) = \sigma_\nu^2$. **Hint:** $E(u'Qu) = E\{\text{tr}(u'Qu)\} = E\{\text{tr}(uu'Q)\} = \text{tr}\{E(uu')Q\} = \text{tr}(\Omega Q)$.

5. *Swamy and Arora (1972) Estimates of the Variance-Components.*

   (a) Show that $\widehat{\widehat{\sigma}}_\nu^2$ given in (12.23) is unbiased for σ_ν^2.

   (b) Show that $\widehat{\widehat{\sigma}}_1^2$ given in (12.26) is unbiased for σ_1^2.

6. *System Estimation.*

   (a) Perform OLS on the system of equations given in (12.27) and show that the resulting estimator is $\widehat{\delta}_{OLS} = (Z'Z)^{-1}Z'y$.

   (b) Perform GLS on this system of equations and show that the resulting estimator is $\widehat{\delta}_{GLS} = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y$ where Ω^{-1} is given in (12.19).

7. *Random Effects Is More Efficient than Fixed Effects.* Using the var$(\widehat{\beta}_{GLS})$ expression below (12.30) and var$(\widetilde{\beta}_{Within}) = \sigma_\nu^2 W_{XX}^{-1}$, show that

$$(\text{var}(\widehat{\beta}_{GLS}))^{-1} - (\text{var}(\widetilde{\beta}_{Within}))^{-1} = \phi^2 B_{XX}/\sigma_\nu^2$$

   which is positive semi-definite. Conclude that var$(\widetilde{\beta}_{Within})-$ var$(\widehat{\beta}_{GLS})$ is positive semi-definite.

8. *Maximum Likelihood Estimation of the Random Effects Model.*

   (a) Using the concentrated likelihood function in (12.34), solve $\partial L_c/\partial\phi^2 = 0$ and verify (12.35).

   (b) Solve $\partial L_c/\partial\beta = 0$ and verify (12.36).

9. *Prediction in the Random Effects Model.*

   (a) For the predictor of $y_{i,T+S}$ given in (12.37), compute $E(u_{i,T+S}u_{it})$ for $t = 1, 2, \ldots, T$ and verify that $w = E(u_{i,T+S}u) = \sigma_\mu^2(\ell_i \otimes \iota_T)$ where ℓ_i is the i-th column of I_N.

   (b) Verify (12.39) by showing that $(\ell_i' \otimes \iota_T')P = (\ell_i' \otimes \iota_T')$.

10. Using the *gasoline demand data* of Baltagi and Griffin (1983), given on the Springer web site as GASOLINE.DAT, reproduce Tables 12.1 through 12.7.

11. *Bounds on s^2 in the Random Effects Model.* For the random one-way error components model given in (12.1) and (12.2), consider the OLS estimator of var$(u_{it}) = \sigma^2$, which is given by $s^2 = e'e/(n - K')$, where e denotes the vector of OLS residuals, $n = NT$ and $K' = K + 1$.

    (a) Show that $E(s^2) = \sigma^2 + \sigma_\mu^2[K' - \text{tr}(I_N \otimes J_T)P_X]/(n - K')$.

    (b) Consider the inequalities given by Kiviet and Krämer (1992) which state that $0 \leq$ mean of $n - K'$ smallest roots of $\Omega \leq E(s^2) \leq$ mean of $n - K'$ largest roots of $\Omega \leq \text{tr}(\Omega)/(n - K')$ where $\Omega = E(uu')$. Show that for the one-way error components model, these bounds are

$$0 \leq \sigma_\nu^2 + (n - TK')\sigma_\mu^2/(n - K') \leq E(s^2) \leq \sigma_\nu^2 + n\sigma_\mu^2/(n - K') \leq n\sigma^2/(n - K').$$

    As $n \to \infty$, both bounds tend to σ^2, and s^2 is asymptotically unbiased, irrespective of the particular evolution of X, see Baltagi and Krämer (1994) for a proof of this result.

12. Verify the relationship between M and M^*, i.e., $MM^* = M^*$, given below (12.47). **Hint:** Use the fact that $Z = Z^*I^*$ with $I^* = (\iota_N \otimes I_{K'})$.

13. Verify that $\dot{M}$ and $\dot{M}^*$ defined below (12.50) are both symmetric, idempotent and satisfy $\dot{M}\dot{M}^* = \dot{M}^*$.

14. For the gasoline data used in problem 10, verify the Chow-test results given below equation (12.51).

15. For the gasoline data, compute the Breusch-Pagan, Honda and Standardized LM tests for H_0; $\sigma_\mu^2 = 0$.

16. If $\widetilde{\beta}$ denotes the LSDV estimator and $\widehat{\beta}_{GLS}$ denotes the GLS estimator, then

    (a) Show that $\widehat{q} = \widehat{\beta}_{GLS} - \widetilde{\beta}$ satisfies $\text{cov}(\widehat{q}, \widehat{\beta}_{GLS}) = 0$.

    (b) Verify equation (12.56).

17. For the gasoline data used in problem 10, replicate the Hausman test results given below equation (12.58).

18. For the cigarette data given as CIGAR.TXT on the Springer web site, reproduce the results given in Table 12.8. See also Baltagi, Griffin and Xiong (2000).

19. *Heteroskedastic Fixed Effects Models.* This is based on Baltagi (1996). Consider the fixed effects model

$$y_{it} = \alpha_i + u_{it} \quad i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T_i$$

    where y_{it} denotes output in industry i at time t and α_i denotes the industry fixed effect. The disturbances u_{it} are assumed to be independent with heteroskedastic variances σ_i^2. Note that the data are unbalanced with different number of observations for each industry.

    (a) Show that OLS and GLS estimates of α_i are identical.

    (b) Let $\sigma^2 = \sum_{i=1}^N T_i\sigma_i^2/n$ where $n = \sum_{i=1}^N T_i$, be the average disturbance variance. Show that the GLS estimator of σ^2 is unbiased, whereas the OLS estimator of σ^2 is biased. Also show that this bias disappears if the data are balanced or the variances are homoskedastic.

    (c) Define $\lambda_i^2 = \sigma_i^2/\sigma^2$ for $i = 1, 2 \ldots, N$. Show that for $\alpha' = (\alpha_1, \alpha_2, \ldots, \alpha_N)$

$$\begin{aligned} & E[\text{estimated var}(\widehat{\alpha}_{OLS}) - \text{ true var}(\widehat{\alpha}_{OLS})] \\ = \ & \sigma^2[(n - \sum_{i=1}^N \lambda_i^2)/(n - N)] \text{ diag } (1/T_i) - \sigma^2 \text{ diag } (\lambda_i^2/T_i) \end{aligned}$$

    This problem shows that in case there are no regressors in the unbalanced panel data model, fixed effects with heteroskedastic disturbances can be estimated by OLS, but one has to correct the standard errors.

20. *The Relative Efficiency of the Between Estimator with Respect to the Within Estimator.* This is based on Baltagi (1999). Consider the simple panel data regression model

$$y_{it} = \alpha + \beta x_{it} + u_{it} \quad i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T \tag{1}$$

    where α and β are scalars. Subtract the mean equation to get rid of the constant

$$y_{it} - \bar{y}_{..} = \beta(x_{it} - \bar{x}_{..}) + u_{it} - \bar{u}_{..}, \tag{2}$$

where $\bar{x}_{..} = \Sigma_{i=1}^{N}\Sigma_{t=1}^{T}x_{it}/NT$ and $\bar{y}_{..}$ and $\bar{u}_{..}$ are similarly defined. Add and subtract $\bar{x}_{i.}$ from the regressor in parentheses and rearrange

$$y_{it} - \bar{y}_{..} = \beta(x_{it} - \bar{x}_{i.}) + \beta(\bar{x}_{i.} - \bar{x}_{..}) + u_{it} - \bar{u}_{..} \tag{3}$$

where $\bar{x}_{i.} = \Sigma_{t=1}^{T}x_{it}/T$. Now run the unrestricted least squares regression

$$y_{it} - \bar{y}_{..} = \beta_w(x_{it} - \bar{x}_{i.}) + \beta_b(\bar{x}_{i.} - \bar{x}_{..}) + u_{it} - \bar{u}_{..} \tag{4}$$

where β_w is not necessarily equal to β_b.

(a) Show that the least squares estimator of β_w from (4) is the Within estimator and that of β_b is the Between estimator.

(b) Show that if $u_{it} = \mu_i + \nu_{it}$ where $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ and $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$ independent of each other and among themselves, then ordinary least squares (OLS) is equivalent to generalized least squares (GLS) on (4).

(c) Show that for model (1), the relative efficiency of the Between estimator with respect to the Within estimator is equal to $(B_{XX}/W_{XX})[(1-\rho)/(T\rho + (1-\rho))]$, where $W_{XX} = \Sigma_{i=1}^{N}\Sigma_{t=1}^{T}(x_{it} - \bar{x}_{i.})^2$ denotes the Within variation and $B_{XX} = T\Sigma_{i=1}^{N}(\bar{x}_{i.} - \bar{x}_{..})^2$ denotes the Between variation. Also, $\rho = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\nu^2)$ denotes the equicorrelation coefficient.

(d) Show that the square of the t-statistic used to test H_0; $\beta_w = \beta_b$ in (4) yields exactly Hausman's (1978) specification test.

21. For the *crime example* of Cornwell and Trumbull (1994) studied in Chapter 11. Use the panel data given as CRIME.DAT on the Springer web site to replicate the Between and Within estimates given in Table 1 of Cornwell and Trumbull (1994). Compute 2SLS and Within-2SLS (2SLS with county dummies) using offense mix and per capita tax revenue as instruments for the probability of arrest and police per capita. Comment on the results.

22. Consider the Arellano and Bond (1991) dynamic employment equation for 140 UK companies over the period 1979–1984. Replicate all the estimation results in Table 4 of Arellano and Bond (1991, p. 290).

References

This chapter is based on Baltagi (2008).

Ahn, S.C. and P. Schmidt (1995), "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68: 5–27.

Amemiya, T. (1971), "The Estimation of the Variances in a Variance-Components Model," *International Economic Review*, 12: 1–13.

Anderson, T.W. and C. Hsiao (1982), "Formulation and Estimation of Dynamic Models Using Panel Data, *Journal of Econometrics*, 18: 47–82.

Arellano, M. (1989), "A Note on the Anderson-Hsiao Estimator for Panel Data," *Economics Letters*, 31: 337–341.

Arellano, M. (1993), "On the Testing of Correlated Effects With Panel Data," *Journal of Econometrics*, 59: 87–97.

Arellano, M. and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and An Application to Employment Equations," *Review of Economic Studies*, 58: 277–297.

Balestra, P. (1973), "Best Quadratic Unbiased Estimators of the Variance-Covariance Matrix in Normal Regression," *Journal of Econometrics*, 2: 17–28.

Baltagi, B.H. (1981), "Pooling: An Experimental Study of Alternative Testing and Estimation Procedures in a Two-Way Errors Components Model," *Journal of Econometrics*, 17: 21–49.

Baltagi, B.H. (1996), "Heteroskedastic Fixed Effects Models," Problem 96.5.1, *Econometric Theory*, 12: 867.

Baltagi, B.H. (1999), "The Relative Efficiency of the Between Estimator with Respect to the Within Estimator," Problem 99.4.3, *Econometric Theory*, 15: 630–631.

Baltagi, B.H. (2008), *Econometric Analysis of Panel Data* (Wiley: Chichester).

Baltagi, B.H. and J.M. Griffin (1983), "Gasoline Demand in the OECD: An Application of Pooling and Testing Procedures," *European Economic Review*, 22: 117–137.

Baltagi, B.H., J.M. Griffin and W. Xiong (2000), "To Pool or Not to Pool: Homogeneous Versus Heterogeneous Estimators Applied to Cigarette Demand," *Review of Economics and Statistics*, 82: 117–126.

Baltagi, B.H. and W. Krämer (1994), "Consistency, Asymptotic Unbiasedness and Bounds on the Bias of s^2 in the Linear Regression Model with Error Components Disturbances," *Statistical Papers*, 35: 323–328.

Breusch, T.S. (1987), "Maximum Likelihood Estimation of Random Effects Models," *Journal of Econometrics*, 36: 383–389.

Breusch, T.S. and A.R. Pagan (1980), "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics," *Review of Economic Studies*, 47: 239–253.

Card (1990), "The Impact of the Mariel Boat Lift on the Miami Labor Market," *Industrial and Labor Relations Review*, 43: 245–253.

Chow, G.C. (1960), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28: 591–605.

Cornwell, C. and W.N. Trumbull (1994), "Estimating the Economic Model of Crime with Panel Data," *Review of Economics and Statistics* 76: 360–366.

Evans, M.A. and M.L. King (1985), "Critical Value Approximations for Tests of Linear Regression Disturbances," *Australian Journal of Statistics*, 27: 68–83.

Fisher, F.M. (1970), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note," *Econometrica*, 38: 361–366.

Fuller, W.A. and G.E. Battese (1974), "Estimation of Linear Models with Cross-Error Structure," *Journal of Econometrics*, 2: 67–78.

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

Graybill, F.A. (1961), *An Introduction to Linear Statistical Models* (McGraw-Hill: New York).

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50: 1029–1054.

Hausman, J.A. (1978), "Specification Tests in Econometrics," *Econometrica*, 46: 1251–1271.

Honda, Y. (1985), "Testing the Error Components Model with Non-Normal Disturbances," *Review of Economic Studies*, 52: 681–690.

Hsiao, C. (2003), *Analysis of Panel Data* (Cambridge University Press: Cambridge).

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepohl and T.C. Lee (1985), *The Theory and Practice of Econometrics* (Wiley: New York).

Kiviet, J.F. and W. Krämer (1992), "Bias of s^2 in the Linear Regression Model with Correlated Errors," *Empirical Economics*, 16: 375–377.

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

Maddala, G.S. and T. Mount (1973), "A Comparative Study of Alternative Estimators for Variance Components Models Used in Econometric Applications," *Journal of the American Statistical Association*, 68: 324–328.

Moulton, B.R. and W.C. Randolph (1989), "Alternative Tests of the Error Components Model," *Econometrica*, 57: 685–693.

Nerlove, M. (1971), "A Note on Error Components Models," *Econometrica*, 39: 383–396.

Nickell, S. (1981), "Biases in Dynamic Models with Fixed Effects,"*Econometrica*, 49: 1417–1426.

Searle, S.R. (1971), *Linear Models* (Wiley: New York).

Sargan, J. (1958), "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26: 393–415.

Swamy, P.A.V.B. and S.S. Arora (1972), "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models," *Econometrica*, 40: 261–275.

Taub, A.J. (1979), "Prediction in the Context of the Variance-Components Model," *Journal of Econometrics*, 10: 103–108.

Taylor, W.E. (1980), "Small Sample Considerations in Estimation from Panel Data," *Journal of Econometrics*, 13: 203–223.

Wallace, T. and A. Hussain (1969), "The Use of Error Components Models in Combining Cross-Section and Time-Series Data," *Econometrica*, 37: 55–72.

Wansbeek, T.J. and A. Kapteyn (1978), "The Separation of Individual Variation and Systematic Change in the Analysis of Panel Data," *Annales de l'INSEE*, 30–31: 659–680.

Wansbeek, T.J. and A. Kapteyn (1982), "A Simple Way to Obtain the Spectral Decomposition of Variance Components Models for Balanced Data," *Communications in Statistics All*, 2105–2112.

Wansbeek, T.J. and A. Kapteyn, (1989), "Estimation of the error components model with incomplete panels," *Journal of Econometrics* 41: 341–361.

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regression and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 57: 348–368.