11.4.5 Time Invariant Variables and Fixed Effects Vector Decomposition

The presence of time invariant variables (TIVs) in the common effects regression presents a vexing problem for the model builder. The significant problem *for the fixed effects model* (FEM) is that the estimator cannot accommodate TIVs. Thus, in the wage equation in Example 11.5, we have omitted three variables of considerable interest from the fixed effects model, *Ed*, *Fem* and *Blk*. If we write the FEM with a set of time invariant variables in it as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

with $\mathbf{Z}$ being the matrix of M TIVs, then the problem becomes one of multicollinearity. Since the columns of matrix $\mathbf{D}$ are a complete set of n dummy variables, any time invariant variable in $\mathbf{Z}$ can be written as a linear combination of the columns of $\mathbf{D}$. Let the mth column of $\mathbf{Z}$ be the TIV, $\mathbf{Z}(m) = (z_{m1},z_{m1},\ldots,z_{m2},z_{m2},\ldots,\ldots,z_{mn},z_{mn},\ldots)'$; each specific value, z_{mi}, is repeated T_i times. Then $\mathbf{Z}(m)$ equals $\mathbf{D}\mathbf{z}_m$ where $\mathbf{z}_m$ is the $n\times 1$ vector $(z_{m1},z_{m2},\ldots,z_{mn})'$. Collecting all M columns, we have $\mathbf{Z} = \mathbf{D}\mathbf{Z}_n$ where $\mathbf{Z}_n$ is the $n\times m$ matrix $(\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_m)$.

If we attempt to compute the LSDV estimator of $(\boldsymbol{\beta}',\boldsymbol{\gamma}')'$ of (11-13) using the transformed variables $\mathbf{M}_\mathbf{D}[\mathbf{X},\mathbf{Z}]$, the columns of $\mathbf{Z}$ are transformed to deviations from group means, which are columns of zeros since $\mathbf{Z}$ is already the period means, and the transformed data matrix becomes $(\mathbf{M}_\mathbf{D}\mathbf{X},\mathbf{0})$ – since $\mathbf{Z}$ is already in the form of group means, deviations from group means are zero. The LSDV regression cannot be computed with TIVs. In theoretical terms, the problem is that $\boldsymbol{\gamma}$ is not identified. No amount of data can disentangle $\boldsymbol{\gamma}$ from $\boldsymbol{\alpha}$. The model would be

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}(\mathbf{Z}_n\boldsymbol{\gamma}) + \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}[\mathbf{Z}_n\boldsymbol{\gamma} + \boldsymbol{\alpha}] + \boldsymbol{\varepsilon}.$$

In the fixed effects case, the identifying restriction is $\boldsymbol{\gamma} = \mathbf{0}$. That is in a fixed effects model, the coefficients on TIVs are not identified in terms of the moments of the data so their coefficients are fixed at zero, so as to identify $\boldsymbol{\alpha}$.

Plümper and Troeger (2007) have proposed a three step procedure that they label **Fixed Effects Vector Decomposition** (FEVD), that suggests a solution to the problem of estimating coefficients on TIVs in a fixed effects model and, at the same time, brings noticeable gains in the efficiency of estimation of the parameters. The three steps are:

Step 1: Linear regression of $\mathbf{y}$ on $\mathbf{X}$ and $\mathbf{D}$ to estimate $\boldsymbol{\alpha}$. That is, compute the LSDV estimator of $\boldsymbol{\beta}$ in (11-13) and use (11-15) to compute estimates of the individual constant terms.

Step 2: Linear regression of the n estimated constant terms, a_i, $i = 1,\ldots,n$, on a constant term and $\mathbf{Z}_n$. From this regression, we compute the n residuals, $\mathbf{h}_n$. We then expand this vector to the full sample length using $\mathbf{h} = \mathbf{D}\mathbf{h}_n$.

Step 3: Linear regression of $\mathbf{y}$ on $[\mathbf{X},(\mathbf{i},\mathbf{Z}),\mathbf{h}]$, where $\mathbf{i}$ is an overall constant term, to estimate $(\boldsymbol{\beta}, \alpha^0, \boldsymbol{\gamma}, \delta)$ in $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \alpha^0 + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{h}\delta + \boldsymbol{\varepsilon}$.

The suggestion produces some interesting algebraic results that will be instructive for the analysis of this chapter. The surprising result that has apparently gone unnoticed in dozens of recent applications of the technique, but not in several recent comments including Breusch, Ward, Nguyen and Kompas (2010), Chatelain and Ralf (2010) and Greene (2010), is that Step 3 simply reproduces the results in Steps 1 and 2, but the covariance matrix computed for the estimator of $\boldsymbol{\beta}$ at Step 3 is not identical and is unambiguously too small. It is instructive to work through a derivation in detail.

We will prove the following results:

FEVD.1 The estimated coefficients on $\mathbf{X}$ at Step 3 are identical to those at Step 1.

FEVD.2 The estimated coefficients on $(\mathbf{i},\mathbf{Z})$ at Step 3 are identical to those at Step 2.

FEVD.3 The estimated coefficient on $\mathbf{h}$ at Step 3 equals 1.0.

FEVD.4 The sum of squared residuals in the regression at Step 3 is identical to that at Step 1.

FEVD.5 The s^2 computed at Step 3 is less than that at Step 1.

FEVD.6 The asymptotic covariance matrix computed for the estimator of $\boldsymbol{\beta}$ at Step 3 is smaller than that at Step 1 (even though the estimates are algebraically identical) because of FEVD.5 and because the matrix used is smaller.

(Note, there are much more compact proofs of these results. The following approaches are used to demonstrate the the tools we have developed in this and the preceding chapters.)

Proofs of results: Write the results of the three least squares regressions as

(Step 1) $\mathbf{y} = \mathbf{X}\mathbf{b}_{\text{LSDV}} + \mathbf{D}\mathbf{a}_{\text{LSDV}} + \mathbf{e}_{\text{LSDV}}$,
(Step 2) $\mathbf{a}_{\text{LSDV}} = \mathbf{W}_n\mathbf{c}_{\text{LSDV}} + \mathbf{h}_n$ where $\mathbf{W}_n = (\mathbf{i}_n,\mathbf{Z}_n)$,
(Step 3) $\mathbf{y} = \mathbf{X}\mathbf{b}_{\text{FEVD}} + \mathbf{W}\mathbf{c}_{\text{FEVD}} + \mathbf{h}d_{\text{FEVD}} + \mathbf{e}_{\text{FEVD}}$, where $\mathbf{W} = (\mathbf{i},\mathbf{Z})$.

Thus, $\mathbf{W}$ at Step 3 includes the M time invariant variables and an overall constant. To begin, we will establish that $\mathbf{e}_{\text{LSDV}} = \mathbf{e}_{\text{FEVD}}$. Recall that $\mathbf{Z} = \mathbf{D}\mathbf{Z}_n$ and $\mathbf{i} = \mathbf{D}\mathbf{i}_n$ where $\mathbf{i}_n$ is an $n\times 1$ column vector of ones. The residuals in (Step 2) are $\mathbf{h}_n = \mathbf{a}_{\text{LSDV}} - \mathbf{W}_n\mathbf{c}_{\text{LSDV}}$ and $\mathbf{h} = \mathbf{D}\mathbf{h}_n$. Therefore, the result at (Step 3) is equivalent to

$$\mathbf{y} = \mathbf{X}\mathbf{b}_{\text{FEVD}} + \mathbf{D}\mathbf{W}_n\mathbf{c}_{\text{FEVD}} + \mathbf{D}(\mathbf{a}_{\text{LSDV}} - \mathbf{W}_n\mathbf{c}_{\text{LSDV}})d_{\text{FEVD}} + \mathbf{e}_{\text{FEVD}}.$$

Rearranging it slightly,

$$\mathbf{y} = \mathbf{X}\mathbf{b}_{\text{FEVD}} + \mathbf{D}\mathbf{a}_{\text{LSDV}} + \mathbf{D}\mathbf{W}_n\mathbf{c}_{\text{FEVD}} - \mathbf{D}\mathbf{W}_n\mathbf{c}_{\text{LSDV}}(d_{\text{FEVD}}) + \mathbf{e}_{\text{FEVD}}. \qquad (*)$$

=(11-27)

The first two terms are the predictions from the linear regression of $\mathbf{y}$ on $\mathbf{X}$ and $\mathbf{D}$ and the third and fourth terms simply add more linear combinations of the columns of $\mathbf{D}$. Since $(\mathbf{X},\mathbf{D})$ has (we have assumed) full column rank, least squares regression (*) must provide the same fit as (Step 1). The residuals must be identical; that is, $\mathbf{e}_{\text{FEVD}} = \mathbf{e}_{\text{LSDV}}$. Now, premultiply (*) by $\mathbf{X}'\mathbf{M}_{\mathbf{D}}$. Since $\mathbf{M}_{\mathbf{D}}\mathbf{D} = \mathbf{0}$ and $\mathbf{M}_{\mathbf{D}}\mathbf{e}_{\text{LSDV}} = \mathbf{e}_{\text{LSDV}}$, we find

$$\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{y} = \mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{X}\mathbf{b}_{\text{FEVD}} + \mathbf{X}'\mathbf{e}_{\text{LSDV}}.$$

Since $\mathbf{X}'\mathbf{e}_{\text{LSDV}} = \mathbf{0}$ from (Step 1), we have $\mathbf{b}_{\text{FEVD}} = (\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{y}) = \mathbf{b}_{\text{LSDV}}$ which proves FEVD.1.

To compute $\mathbf{c}_{\text{FEVD}}$ at (Step 3) we have at the solution (using $\mathbf{b}_{\text{FEVD}} = \mathbf{b}_{\text{LSDV}}$ and $\mathbf{e}_{\text{FEVD}} = \mathbf{e}_{\text{LSDV}}$)

$$\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{LSDV}} = \mathbf{W}\mathbf{c}_{\text{FEVD}} + \mathbf{h}d_{\text{FEVD}} + \mathbf{e}_{\text{LSDV}}.$$

Premultiply this expression by $\mathbf{W}'$. From (Step 2) $\mathbf{W}'\mathbf{h} = \mathbf{W}_n{}'\mathbf{D}'\mathbf{D}\mathbf{h}_n = \mathbf{0}$. This is true because $\mathbf{D}'\mathbf{D}$ is a diagonal matrix with T_i on the diagonals. Thus, each element in $\mathbf{W}'\mathbf{h}$ is $T_i\mathbf{W}(m)'\mathbf{h}_n = 0$, where $\mathbf{W}(m)$ is the mth column of $\mathbf{W}_n$. From (Step 3), $\mathbf{W}'\mathbf{e}_{\text{FEVD}} = \mathbf{W}'\mathbf{e}_{\text{LSDV}} = \mathbf{0}$. Thus,

$$\mathbf{W}'(\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{LSDV}}) = \mathbf{W}'\mathbf{W}\mathbf{c}_{\text{FEVD}}$$

so

$$\mathbf{c}_{\text{FEVD}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'(\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{LSDV}}).$$

From (Step 1), $\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{LSDV}} = \mathbf{D}\mathbf{a}_{\text{LSDV}} + \mathbf{e}_{\text{LSDV}}$. Since $\mathbf{W}'\mathbf{e}_{\text{FEVD}} = \mathbf{W}'\mathbf{e}_{\text{LSDV}} = \mathbf{0}$, from (Step 3),

$$\mathbf{c}_{\text{FEVD}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{D}\mathbf{a}_{\text{LSDV}}.$$

But, by premultiplying (Step 2) by $\mathbf{D}$, we find $\mathbf{D}\mathbf{a}_{\text{LSDV}} = \mathbf{D}\mathbf{W}_n\mathbf{c}_{\text{LSDV}} + \mathbf{D}\mathbf{h}_n$. It follows that the solution is

minus

$$\mathbf{c}_{\text{LSDV}} = (\mathbf{W}_n'\mathbf{D}'\mathbf{D}\mathbf{W}_n)^{-1}\mathbf{W}_n'\mathbf{D}'\mathbf{D}\mathbf{a}_{\text{LSDV}} + (\mathbf{W}_n'\mathbf{D}'\mathbf{D}\mathbf{W}_n)^{-1}\mathbf{W}_n'\mathbf{D}'\mathbf{D}\mathbf{h}_n.$$

The second term is zero as shown earlier. The end result is $\mathbf{c}_{\text{LSDV}} = \mathbf{c}_{\text{FEVD}}$ which is FEVD.2.

Once again using (Step 3), we now solve for d_{FEVD} using what we already have. The solution is in

$$\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{LSDV}} - \mathbf{W}\mathbf{c}_{\text{LSDV}} = \mathbf{h}d_{\text{FEVD}} + \mathbf{e}_{\text{LSDV}}.$$

minus

But, $\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{LSDV}} = \mathbf{a} + \mathbf{e}_{\text{LSDV}} = \mathbf{D}\mathbf{a}_{\text{LSDV}} + \mathbf{e}_{\text{LSDV}}$ and $\mathbf{W}\mathbf{c}_{\text{LSDV}} = \mathbf{a} - \mathbf{h} = \mathbf{D}\mathbf{a}_{\text{LSDV}} - \mathbf{h}$. Inserting these,

$$\mathbf{D}\mathbf{a}_{\text{LSDV}} + \mathbf{e}_{\text{LSDV}} - \mathbf{D}\mathbf{a}_{\text{LSDV}} + \mathbf{h} = \mathbf{h}d_{\text{FEVD}} + \mathbf{e}_{\text{LSDV}}$$

or

$$\mathbf{h} + \mathbf{e}_{\text{LSDV}} = \mathbf{h}d_{\text{FEVD}} + \mathbf{e}_{\text{LSDV}},$$

from which it follows that $d_{\text{FEVD}} = 1$. This proves FEVD.3.

FEVD.4 has already been shown since $\mathbf{e}_{\text{FEVD}} = \mathbf{e}_{\text{LSDV}}$. The R^2s in the two regressions are the same as well, as $R_{\text{FEVD}}^2 = 1 - (\mathbf{e}_{\text{FEVD}}'\mathbf{e}_{\text{FEVD}}/\mathbf{y}'\mathbf{M}^0\mathbf{y}) = R_{\text{LSDV}}^2$ since the residual vectors are identical. [See (3-26).] But,

$$s_{\text{FEVD}}^2 = \mathbf{e}_{\text{FEVD}}'\mathbf{e}_{\text{FEVD}}/(\Sigma T_i - K - M - 1 - 1) < s_{\text{LSDV}}^2 = \mathbf{e}_{\text{LSDV}}'\mathbf{e}_{\text{LSDV}}/(\Sigma_i T_i - K - n).$$

The difference is the degrees of freedom correction, which can be large. In our example to follow, $DF_{\text{FEVD}} = 4165 - 9 - 3 - 1 - 1 = 4151$ while $DF_{\text{LSDV}} = 4165 - 9 - 595 = 3561$. For the example, then, $s_{\text{FEVD}}^2 / s_{\text{LSDV}}^2 = 0.85787$. This establishes FEVD.5.

To establish FEVD.6, based on (11-16), we are going to compare (11-17)

minus

$$\text{Est.Asy.Var}[\mathbf{b}_{\text{FEVD}}] = s_{\text{FEVD}}^2(\mathbf{X}'\mathbf{M}_{\mathbf{W},\mathbf{h}}\mathbf{X})^{-1}$$

to

minus

$$\text{Est.Asy.Var}[\mathbf{b}_{\text{LSDV}}] = s_{\text{LSDV}}^2(\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{X})^{-1}.$$

We have already established that $s_{\text{LSDV}}^2 > s_{\text{FEVD}}^2$. To compare the matrices, we will compare their inverses, and show that the difference matrix

$$\mathbf{A} = \mathbf{X}'\mathbf{M}_{\mathbf{W},\mathbf{h}}\mathbf{X} - \mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{X}$$

is positive definite. This will imply that the inverse matrix in Est.Asy.Var[$\mathbf{b}_{\text{FEVD}}$] is smaller than that in Est.Asy.Var[$\mathbf{b}_{\text{LSDV}}$]. To show this, we note that $(\mathbf{W},\mathbf{h}) = \mathbf{D}(\mathbf{W}_n,\mathbf{h}_n)$ is $M+2$ linear combinations of the columns of $\mathbf{D}$ while $\mathbf{D}$ is all n columns of $\mathbf{D}$. For convenience, let $\mathbf{R} = (\mathbf{W},\mathbf{h})$. The residuals defined by $\mathbf{M}_{\mathbf{D}}\mathbf{X}$ [see (3-15)] are obtained by regressions of $\mathbf{X}$ on all n columns of $\mathbf{D}$. They will be identical to the residuals obtained by regression of $\mathbf{X}$ on any n linearly independent combinations of the columns of $\mathbf{D}$. For these, we will use $[\mathbf{R},\mathbf{Q}]$ where $\mathbf{Q}$ is orthogonal to $\mathbf{R}$. Therefore $\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{X} = \mathbf{X}'\mathbf{M}_{\mathbf{R},\mathbf{Q}}\mathbf{X}$. Expanding this, we have

minus

$$\mathbf{A} = \mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{R}(\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'\mathbf{X} - \mathbf{X}'\mathbf{X} + \mathbf{X}'(\mathbf{R} \quad \mathbf{Q})\left[\begin{pmatrix}\mathbf{R}'\\ \mathbf{Q}'\end{pmatrix}(\mathbf{R} \quad \mathbf{Q})\right]^{-1}\begin{pmatrix}\mathbf{R}'\\ \mathbf{Q}'\end{pmatrix}\mathbf{X}.$$

The inverse matrix is simplified by $\mathbf{R'Q} = \mathbf{0}$, so the bracketed matrix and its inverse are block diagonal. Multiplying it out, we find

minus minus

$$\mathbf{A} = \mathbf{X'Q(Q'Q)^{-1}Q'X} = \mathbf{X'(I-M_Q)X}.$$

minus minus

Since $\mathbf{I}-\mathbf{M_Q}$ is idempotent, $\mathbf{A} = \mathbf{X'(I-M_Q)'(I-M_Q)X} = \mathbf{X^{*\prime}X^*}$ is positive definite. This establishes that the computed covariance matrix for $\mathbf{b}_{FEVD}$ will always be strictly smaller than that for $\mathbf{b}_{LSDV}$, which is FEVD.6.

This leaves what should appear to be a loose end in the analysis. How was it possible to estimate $\boldsymbol{\gamma}$ in (Step 2 or Step 3) given that it is unidentified in the original model? The answer is the crucial assumption noted at several points in the preceding. From the original specification $\mathbf{Z}$ is uncorrelated with $\boldsymbol{\varepsilon}$. But, for the regression in (Step 2) to estimate a nonzero $\boldsymbol{\gamma}$, *it must be further assumed that* $\mathbf{z}_i$ *is uncorrelated with* u_i. This restricts the original fixed effects model – it is a hybrid in which the time varying variables are allowed to be correlated with u_i but the time invariant variables are not. The authors note this on page 6 and in their footnote 7 where they state "If the time invariant variables are assumed to be orthogonal to the unobserved unit effects – i.e., if the assumption underlying our estimator is correct – the estimator is consistent. If this assumption is violated, the estimated coefficients for the time-invariant variables are biased.... Note that the estimated coefficients of the time-varying variables remain unbiased even in the presence of correlated unit effects. However, the assumption underlying a FE model must be satisfied (*no correlated time-varying variables may exist*)." (Emphasis added – it seems that "varying" should be "invariant.") There are other estimators that would consistently $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in this revised model, including the Hausman and Taylor estimator discussed in Section 11.8.1 and instrumental variables estimators suggested by Breusch et al. (2010) and by Chatelain and Ralf (2010).

umlaut

The problem of primary interest in Plümper and Troeger was an intermediate case somewhat different from what we have examined here. There are two directions of the work. If only some of the elements of $\mathbf{Z}$ but not all of them, are correlated with u_i, then we obtain the setting analysed by Hausman and Taylor that is examined in Section 11.8.1. Plümper and Troeger's FEVD estimator will, in that instance be an inconsistent estimator that may have a smaller variance than the IV estimator proposed by Hausman and Taylor. The second case the authors are interested in is when $\mathbf{Z}$ is not strictly time invariant, but is "slowly changing." When there is very little within groups variation, for example, as shown for the World Health Organization data in Example 11.4, then, once again, the estimator suggested here may have some advantages over instrumental variables and other treatments. In that case, when there are no strictly time invariant variables in the model, then the analysis is governed by the random effects model discussed in the next section.

previously

AQ: Please clarify which "authors" you are referring to.

be

AQ: Provide year for Plümper & Troeger?

umlaut

TB 11.5

Example 11.5 Fixed Effects Wage Equation

Table 11.5 presents the estimated wage equation with individual effects for the Cornwell and Rupert data used in Examples 11.1 and 11.3. The model includes three time-invariant variables, *Ed*, *Fem*, *Blk*, that must be dropped from the equation. As a consequence, the fixed effects estimates computed here are not comparable to the results for the pooled model already examined. For comparison, the least squares estimates with panel robust standard errors are also presented. We have also added a set of time dummy variables to the model. The *F* statistic for testing the significance of the individual effects based on the R2s for the equations is

R2's

minus

$$F[594,3561] = \frac{(0.9072422 - 0.3154548)/594}{(1 - 0.9072422)/(4165 - 9 - 595)} = 38.247$$

minus

The critical value for the *F* table with 594 and 3561 degrees of freedom is 1.106, so the evidence is strongly in favor of an individual-specific effect. As often happens, the fit of the model increases greatly when the individual effects are added. We have also added time effects to the model. The model with time effects without the individual effects are in the second column results. The *F* statistic for testing the significance of the time effects (in the absence of the individual effects) is

15

minus

$$F[6,4149] = \frac{(0.4636788 - 0.3154548)/6}{(1 - 0.4636788)/(4165 - 10 - 6)} = 191.11,$$

minus

The critical value from the *F* table is 2.101, so the hypothesis that the time effects are zero is also rejected. The last column of results shows the model with both time and individual effects. For this model it is necessary to drop a second time effect because the experience variable, *Exp*, is an individual specific time trend. The *Exp* variable can be expressed as

$$Exp_{i,t} = E_{i,0} + (t - 1),\ t = 1, \ldots, 7,$$

which can be expressed as a linear combination of the individual dummy variable and the six time variables. For the last model, we have dropped the first and last of the time effects. In this model, the *F* statistic for testing the significance of the time effects is

minus

$$F[5,3556] = \frac{(0.9080847 - 0.9072422)/5}{(1 - 0.9080847)/(4165 - 9 - 5 - 5595)} = 6.519.$$

minus

The time effects remain significant—the critical value is 2.217—but the test statistic is considerably reduced. The time effects reveal a striking pattern. In the equation without the individual effects, we find a steady increase in wages of 7–9 percent per year. But, when the individual effects are added to the model, this progression disappears.

It might seem appropriate to compute the robust standard errors for the fixed effects estimator as well as for the pooled estimator. However, in principle, that should be unnecessary. If the model is correct and completely specified, then the individual effects should be capturing the omitted heterogeneity, and what remains is a classical, homoscedastic, nonautocorrelated disturbance. This does suggest a rough indicator of the appropriateness of the model specification. If the conventional asymptotic covariance matrix in (11-16) and the robust estimator in (11-3), with $\mathbf{X}_i$ replaced with the data in group mean deviations form, give very different estimates, one might question the model specification. [This is the logic that underlies White's (1982a) information matrix test (and the extensions by Newey (1985a) and Tauchen (1985).] The robust standard errors are shown in parentheses under those for the fixed effects estimates in the sixth column of Table 11.5. They are considerably higher than the uncorrected standard errors—50

317

percent to 100 percent—which might suggest that the fixed effects specification should be reconsidered.

The FEVD computations are shown in Table 11.5 as well. The third set of results, marked "Individual Effects," shows the (Step 1) and (Step 2) results. Note that these are computed in two least squares regressions. The second step is indicated by the heavy box. The fit measures are not shown for this intermediate step. The (Step 3) results are shown in the last two columns of the table. As anticipated, the estimated coefficients match the first and second step regressions. For $\mathbf{b}_{LSDV}$, the standard errors have fallen by a factor of 2 to 4. For $\mathbf{c}_{LSDV}$, the estimators of γ, they have fallen by a factor of 7 to 10. In view of the previous analytic results, the estimates in the last column of Table 11.5 would be viewed as overly optimistic.

11-30

Note minus signs

Table 11.5 Fixed Effects Estimates of the Cornwell and Rupert Wage Equation

Variable	Pooled		Time Effects		Individual Effects		Time and Ind. Effects		FEVD Step 3	
	Estimate	Std.Error*	Estimate	Std.Error*	Estimate	Std.Error (Robust)	Estimate	Std.Err	Estimate	Standard Error
Constant	5.8802	0.09654	5.6963	0.09425						
Exp	0.03611	0.0045241	0.02738	0.004556	0.1132	0.002471 (0.00437)	0.1114	0.002618	0.1132	0.00100
Exp^2	-0.00066	0.0001013	-0.00053	0.000101	-0.00042	0.000055 (0.000089)	-0.00004	0.000054	-0.00042	0.0000192
Wks	0.00446	0.001725	0.00409	0.001694	0.00084	0.000600 (0.00094)	0.00068	0.0005991	0.00084	0.00044
Occ	-0.3176	0.02721	-0.3045	0.02684	-0.02148	0.01378 (0.02052)	-0.01916	0.01275	-0.02148	0.00596
Ind	0.03213	0.02521	0.04010	0.02489	0.01921	0.01545 (0.02450)	0.02076	0.1540	0.01921	0.00476
South	-0.1137	0.028626	-0.1157	0.02834	-0.00186	0.03430 (0.09646)	0.00309	0.03419	-0.00186	0.00506
SMSA	0.1586	0.025967	0.1722	0.02566	-0.04247	0.01942 (0.03185)	-0.04188	0.01937	-0.04247	0.00504
MS	0.3203	0.03487	0.3425	0.03459	-0.02973	0.01898 (0.02902)	-0.02856	0.018918	-0.02973	0.00831
Union	0.06975	0.026618	0.06272	0.02578	0.03278	0.01492 (0.02708)	0.02952	0.01488	0.03278	0.00517
Constant					2.8286	0.18599			2.8286	0.03315
Fem					-0.13003	0.12557			-0.13003	0.01024
Ed					0.14438	0.01403			0.14438	0.00121
Blk					-0.27507	0.15440			-0.14438	0.00891
hi									1.00000	0.00683
Year 1			0.0000	0.0000			0.0000	0.0000		
Year 2			0.07812	0.006860			-0.00775	0.008167		
Year 3			0.2050	0.01072			0.02557	0.007769		
Year 4			0.2926	0.01125			0.02845	0.007639		
Year 5			0.3724	0.01095			0.02418	0.007772		
Year 6			0.4498	0.01245			0.00737	0.008161		
Year 7			0.5422	0.013015			0.0000	0.0000		
e'e	607.1265		475.6659		82.26732		81.52012		82.26732	
Deg.Free	4155		4149		3561		3557		4151	
s	0.3822588		0.3385940		0.1519944		0.1514089		0.1407788	
R^2	0.3154548		0.4636788		0.9072422		0.9080847		0.9072422	

*Robust standard errors using (11-3) including finite population correction $[(\Sigma_i T_i)-1] / [(\Sigma_i T_i) - K-m] \times n/(n-1)$.

minus minus

11-31

column results. The F statistic for testing the significance of the time effects (in the absence of the individual effects) is

$$F[6,4149] = \frac{(0.4636788 - 0.3154548)/6}{(1 - 0.4636788)/(4165 - 10 - 6)} = 191.11.$$

The critical value from the F table is 2.101, so the hypothesis that the time effects are zero is also rejected. The last column of results shows the model with both time and individual effects. For this model it is necessary to drop a second time effect because the experience variable, *Exp*, is an individual specific time trend. The *Exp* variable can be expressed as

$$Exp_{i,t} = E_{i,0} + (t - 1), t = 1, \ldots, 7.$$

which can be expressed as a linear combination of the individual dummy variable and the six time variables. For the last model, we have dropped the first and last of the time effects. In this model, the F statistic for testing the significance of the time effects is

$$F[5,3556] = \frac{(0.9080847 - 0.9072422)/5}{(1 - 0.9080847)/(4165 - 9 - 5 - 595)} = 6.519.$$

The time effects remain significant—the critical value is 2.217—, but the test statistic is considerably reduced. The time effects reveal a striking pattern. In the equation without the individual effects, we find a steady increase in wages of 7–8 percent per year. But, when the individual effects are added to the model, this progression disappears.

It might seem appropriate to compute the robust standard errors for the fixed effects estimator as well as for the pooled estimator. However, in principle, that should be unnecessary. If the model is correct and completely specified, then the individual effects should be capturing the omitted heterogeneity, and what remains is a classical, homoscedastic, nonautocorrelated disturbance. This does suggest a rough indicator of the appropriateness of the model specification. If the conventional asymptotic covariance matrix in (~~9~~-16) and the 11 robust estimator in (~~9~~-3), with $\mathbf{X}_i$ replaced with the data in group mean deviations form, give very different estimates, one might question the model specification. [This is the logic that underlies White's (1982a) information matrix test (and the extensions by Newey (1985a) and Tauchen (1985).] The robust standard errors are shown in parentheses under those for the fixed effects estimates in the sixth column of Table ~~9.4.~~ They are considerably higher than the uncorrected standard errors—50 percent to 100 percent—which might suggest that the fixed effects specification should be reconsidered.

11.5

11 ~~9~~.5 RANDOM EFFECTS

The fixed effects model allows the unobserved individual effects to be correlated with the included variables. We then modeled the differences between units strictly as parametric shifts of the regression function. This model might be viewed as applying only to the cross-sectional units in the study, not to additional ones outside the sample. For example, an intercountry comparison may well include the full set of countries for which it is reasonable to assume that the model is constant. If the individual effects are strictly uncorrelated with the regressors, then it might be appropriate to model the individual specific constant terms as randomly distributed across cross-sectional units. This view would be appropriate if we believed that sampled cross-sectional units were drawn from a large population. It would certainly be the case for the longitudinal data sets listed in the introduction to this chapter.[11] The payoff to this form is that it greatly reduces

FN 11

[11] This distinction is not hard and fast; it is purely heuristic. We shall return to this issue later. See Mundlak (1978) for methodological discussion of the distinction between fixed and random effects.

the number of parameters to be estimated. The cost is the possibility of inconsistent estimates, should the assumption turn out to be inappropriate.

Consider, then, a reformulation of the model

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + (\alpha + u_i) + \varepsilon_{it}, \qquad 11\text{-}28\ (\cancel{9}\text{-}25)$$

where there are K regressors including a constant and now the single constant term is the mean of the unobserved heterogeneity, $E[\mathbf{z}_i'\boldsymbol{\alpha}]$. The component u_i is the random heterogeneity specific to the ith observation and is constant through time; recall from
11 Section ~~9~~.2.1, $u_i = \{\mathbf{z}_i'\boldsymbol{\alpha} - E[\mathbf{z}_i'\boldsymbol{\alpha}]\}$. For example, in an analysis of families, we can view u_i as the collection of factors, $\mathbf{z}_i'\boldsymbol{\alpha}$, not in the regression that are specific to that family. We continue to assume strict exogeneity:

$$\begin{aligned}
E[\varepsilon_{it}\,|\,\mathbf{X}] &= E[u_i\,|\,\mathbf{X}] = 0,\\
E[\varepsilon_{it}^2\,|\,\mathbf{X}] &= \sigma_\varepsilon^2,\\
E[u_i^2\,|\,\mathbf{X}] &= \sigma_u^2, \qquad\qquad 11\text{-}29\ (\cancel{9}\text{-}26)\\
E[\varepsilon_{it}u_j\,|\,\mathbf{X}] &= 0 \quad \text{for all } i, t, \text{ and } j,\\
E[\varepsilon_{it}\varepsilon_{js}\,|\,\mathbf{X}] &= 0 \quad \text{if } t \neq s \text{ or } i \neq j,\\
E[u_iu_j\,|\,\mathbf{X}] &= 0 \quad \text{if } i \neq j.
\end{aligned}$$

As before, it is useful to view the formulation of the model in blocks of T observations for group i, $\mathbf{y}_i$, $\mathbf{X}_i$, $u_i\mathbf{i}$, and $\boldsymbol{\varepsilon}_i$. For these T observations, let

$$\eta_{it} = \varepsilon_{it} + u_i$$

and

$$\boldsymbol{\eta}_i = [\eta_{i1}, \eta_{i2}, \ldots, \eta_{iT}]'.$$

KT

In view of this form of η_{it}, we have what is often called an **error components model.** For this model,

$$\begin{aligned}
E[\eta_{it}^2\,|\,\mathbf{X}] &= \sigma_\varepsilon^2 + \sigma_u^2,\\
E[\eta_{it}\eta_{is}\,|\,\mathbf{X}] &= \sigma_u^2, \quad t \neq s \qquad\qquad 11\text{-}30\ (\cancel{9}\text{-}27)\\
E[\eta_{it}\eta_{js}\,|\,\mathbf{X}] &= 0 \qquad \text{for all } t \text{ and } s \text{ if } i \neq j.
\end{aligned}$$

For the T observations for unit i, let $\boldsymbol{\Sigma} = E[\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\,|\,\mathbf{X}]$. Then

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_\varepsilon^2 + \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ & \cdots & & & \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_\varepsilon^2 + \sigma_u^2 \end{bmatrix} = \sigma_\varepsilon^2\mathbf{I}_T + \sigma_u^2\mathbf{i}_T\mathbf{i}_T', \qquad 11\text{-}31\ (\cancel{9}\text{-}28)$$

where $\mathbf{i}_T$ is a $T \times 1$ column vector of 1s. Because observations i and j are independent, the disturbance covariance matrix for the full nT observations is

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \mathbf{0} & \cdots & \mathbf{0} \\ & & & \vdots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma} \end{bmatrix} = \mathbf{I}_n \otimes \boldsymbol{\Sigma}. \qquad \text{11-32 } (\cancel{9}\text{-}29)$$

11 ~~9~~.5.1 GENERALIZED LEAST SQUARES

The generalized least squares estimator of the slope parameters is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} = \left(\sum_{i=1}^{n} \mathbf{X}_i'\boldsymbol{\Sigma}^{-1}\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{n} \mathbf{X}_i'\boldsymbol{\Sigma}^{-1}\mathbf{y}_i\right).$$

To compute this estimator as we did in Chapter 9 ~~8~~ by transforming the data and using ordinary least squares with the transformed data, we will require $\boldsymbol{\Omega}^{-1/2} = [\mathbf{I}_n \otimes \boldsymbol{\Sigma}]^{-1/2}$. We need only find $\boldsymbol{\Sigma}^{-1/2}$, which is

$$\boldsymbol{\Sigma}^{-1/2} = \frac{1}{\sigma_\varepsilon}\left[\mathbf{I} - \frac{\theta}{T}\mathbf{i}_T\mathbf{i}_T'\right],$$

where

$$\theta = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T\sigma_u^2}}.$$

The transformation of $\mathbf{y}_i$ and $\mathbf{X}_i$ for GLS is therefore

$$\boldsymbol{\Sigma}^{-1/2}\mathbf{y}_i = \frac{1}{\sigma_\varepsilon}\begin{bmatrix} y_{i1} - \theta\overline{y}_{i.} \\ y_{i2} - \theta\overline{y}_{i.} \\ \vdots \\ y_{iT} - \theta\overline{y}_{i.} \end{bmatrix}, \qquad \text{11 } (\cancel{9}\text{-}30)$$

and likewise for the rows of $\mathbf{X}_i$.[12] For the data set as a whole, then, generalized least squares is computed by the regression of these partial deviations of y_{it} on the same transformations of $\mathbf{x}_{it}$. Note the similarity of this procedure to the computation in the LSDV model, which uses $\theta = 1$ in (11 ~~9~~-14). (One could interpret θ as the effect that would remain if σ_ε were zero, because the only effect would then be u_i. In this case, the fixed and random effects models would be indistinguishable, so this result makes sense.)

It can be shown that the GLS estimator is, like the pooled OLS estimator, a matrix weighted average of the within- and between-units estimators:

$$\hat{\boldsymbol{\beta}} = \hat{\mathbf{F}}^{within}\mathbf{b}^{within} + (\mathbf{I} - \hat{\mathbf{F}}^{within})\mathbf{b}^{between},^{13} \qquad \text{11 } (\cancel{9}\text{-}31)$$

[12]This transformation is a special case of the more general treatment in Nerlove (1971b).

[13]An alternative form of this expression, in which the weighing matrices are proportional to the covariance matrices of the two estimators, is given by Judge et al. (1985).

weighting

11.5.1 LEAST SQUARES ESTIMATION

The model defined by (11-25),

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

with the strict exogeneity assumptions in (11-26) and the covariance matrix detailed in (11-28) and (11-29) is a generalized regression model that fits into the framework we developed in Chapter 9. The disturbances are autocorrelated in that observations are correlated across time within a group, though not across groups. All the implications of Section 9.2.1 would apply here. In particular, the parameters of the random effects model can be estimated consistently, albeit not efficiently, by ordinary least squares (OLS). An appropriate robust asymptotic covariance matrix for the OLS estimator would be given by (11-3).

There are other consistent estimators available as well. By taking deviations from group means, we obtain

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + \varepsilon_{it} - \bar{\varepsilon}_i.$$

This implies that (assuming there are no time invariant regressors in $\mathbf{x}_{it}$), the LSDV estimator of (11-13) is a consistent estimator of $\boldsymbol{\beta}$. (Note that alone among the four estimators to be suggested here, the LSDV estimator is robust to whether the correct specification is actually a random or a fixed model.) As is OLS, LSDV is inefficient since, as we will show below in Section 11.5.2, there is an efficient GLS estimator that is not equal to $\mathbf{b}_{\text{LSDV}}$. The group means (between groups) regression model

$$\bar{y}_i = \alpha + \bar{\mathbf{x}}_{it}'\boldsymbol{\beta} + u_i + \bar{\varepsilon}_i, i = 1,\ldots,n$$

provides a third method of consistently estimating the coefficients $\boldsymbol{\beta}$. None of these is the preferred estimator in this setting, since the GLS estimator will be more efficient than any of them. However, as we saw in Chapters 9 and 10, many generalized regression models are estimated in two steps, with the first step being a robust least squares regression that is used to produce a first round estimate of the variance parameters of the model. That would be the case here as well. To suggest where this logic will lead in Section 11.5.3, note that for the three cases noted, the mean squared residuals would produce the following consistent estimators of functions of the variances:

(Pooled)	plim $[\mathbf{e}_{\text{pooled}}'\mathbf{e}_{\text{pooled}}/(nT)]$	=	$\sigma_u^2 + \sigma_\varepsilon^2$,
(LSDV)	plim $[\mathbf{e}_{\text{LSDV}}'\mathbf{e}_{\text{LSDV}}/(nT)]$	=	$\sigma_\varepsilon^2[1 - 1/T]$,
(Means)	plim $[\mathbf{e}_{\text{means}}'\mathbf{e}_{\text{means}}/(nT)]$	=	$\sigma_u^2 + \sigma_\varepsilon^2/T$.

Any pair of these estimators would provide a two equation method of moments estimator of $(\sigma_u^2, \sigma_\varepsilon^2)$. With these in mind, we will now develop an efficient generalized least squares estimator.

~~where $\mathbf{i}_T$ is a $T \times 1$ column vector of 1s. Because observations i and j are independent, the disturbance covariance matrix for the full nT observations is~~

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \mathbf{0} & \cdots & \mathbf{0} \\ & & & \vdots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma} \end{bmatrix} = \mathbf{I}_n \otimes \boldsymbol{\Sigma}. \quad 11\ (\cancel{9}\text{-}29)$$

11 2 ~~9~~.5.~~1~~ GENERALIZED LEAST SQUARES

The generalized least squares estimator of the slope parameters is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} = \left(\sum_{i=1}^{n} \mathbf{X}_i'\boldsymbol{\Sigma}^{-1}\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{n} \mathbf{X}_i'\boldsymbol{\Sigma}^{-1}\mathbf{y}_i\right).$$

To compute this estimator as we did in Chapter ~~8~~ 9 by transforming the data and using ordinary least squares with the transformed data, we will require $\boldsymbol{\Omega}^{-1/2} = [\mathbf{I}_n \otimes \boldsymbol{\Sigma}]^{-1/2}$. We need only find $\boldsymbol{\Sigma}^{-1/2}$, which is

$$\boldsymbol{\Sigma}^{-1/2} = \frac{1}{\sigma_\varepsilon}\left[\mathbf{I} - \frac{\theta}{T}\mathbf{i}_T\mathbf{i}_T'\right],$$

where

$$\theta = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T\sigma_u^2}}.$$

The transformation of $\mathbf{y}_i$ and $\mathbf{X}_i$ for GLS is therefore

$$\boldsymbol{\Sigma}^{-1/2}\mathbf{y}_i = \frac{1}{\sigma_\varepsilon}\begin{bmatrix} y_{i1} - \theta\overline{y}_{i.} \\ y_{i2} - \theta\overline{y}_{i.} \\ \vdots \\ y_{iT} - \theta\overline{y}_{i.} \end{bmatrix}, \quad 11\text{-}33\ (\cancel{9}\text{-}30)$$

FN 12

and likewise for the rows of $\mathbf{X}_i$.[12] For the data set as a whole, then, generalized least squares is computed by the regression of these partial deviations of y_{it} on the same transformations of $\mathbf{x}_{it}$. Note the similarity of this procedure to the computation in the LSDV model, which uses $\theta = 1$ in (~~9~~-14) 11-15. (One could interpret θ as the effect that would remain if σ_ε were zero, because the only effect would then be u_i. In this case, the fixed and random effects models would be indistinguishable, so this result makes sense.)

It can be shown that the GLS estimator is, like the pooled OLS estimator, a matrix weighted average of the within- and between-units estimators:

FN 13

$$\hat{\boldsymbol{\beta}} = \hat{\mathbf{F}}^{within}\mathbf{b}^{within} + (\mathbf{I} - \hat{\mathbf{F}}^{within})\mathbf{b}^{between},^{13} \quad 11\text{-}34\ (\cancel{9}\text{-}31)$$

AU: Confirm superscripts are OK set italic

[12] This transformation is a special case of the more general treatment in Nerlove (1971b).

[13] An alternative form of this expression, in which the ~~weighing~~ matrices are proportional to the covariance matrices of the two estimators, is given by Judge et al. (1985).

weighting

11-36

where now,

$$\hat{\mathbf{F}}^{within} = [\mathbf{S}_{xx}^{within} + \lambda \mathbf{S}_{xx}^{between}]^{-1}\mathbf{S}_{xx}^{within},$$

$$\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2} = (1-\theta)^2.$$

To the extent that λ differs from one, we see that the inefficiency of ordinary least squares will follow from an inefficient weighting of the two estimators. Compared with generalized least squares, ordinary least squares places too much weight on the between-units variation. It includes it all in the variation in $\mathbf{X}$, rather than apportioning some of it to random variation across groups attributable to the variation in u_i across units.

Unbalanced panels add a layer of difficulty in the random effects model. The first problem can be seen in (~~9~~-29) 11-32. The matrix $\boldsymbol{\Omega}$ is no longer $\mathbf{I}_n \otimes \boldsymbol{\Sigma}$ because the diagonal blocks in $\boldsymbol{\Omega}$ are of different sizes. There is also groupwise heteroscedasticity in (~~9-30~~) (11-33), because the ith diagonal block in $\boldsymbol{\Omega}^{-1/2}$ is

$$\boldsymbol{\Sigma}_i^{-1/2} = \mathbf{I}_{T_i} - \frac{\theta_i}{T_i}\mathbf{i}_{T_i}\mathbf{i}_{T_i}', \qquad \theta_i = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T_i\sigma_u^2}}.$$

In principle, estimation is still straightforward, because the source of the groupwise heteroscedasticity is only the unequal group sizes. Thus, for GLS, or FGLS with estimated variance components, it is necessary only to use the group-specific θ_i in the transformation in (~~9-30~~) (11-33).

~~9~~ 11.5.~~2~~ 3 FEASIBLE GENERALIZED LEAST SQUARES WHEN Σ IS UNKNOWN

If the variance components are known, generalized least squares can be computed as shown earlier. Of course, this is unlikely, so as usual, we must first estimate the disturbance variances and then use an FGLS procedure. A heuristic approach to estimation of the variance components is as follows:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha + \varepsilon_{it} + u_i \qquad (\text{~~9~~-32}) \; 11\text{-}35$$

and

$$\bar{y}_{i.} = \bar{\mathbf{x}}_{i.}'\boldsymbol{\beta} + \alpha + \bar{\varepsilon}_{i.} + u_i.$$

Therefore, taking deviations from the group means removes the heterogeneity:

$$y_{it} - \bar{y}_{i.} = [\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.}]'\boldsymbol{\beta} + [\varepsilon_{it} - \bar{\varepsilon}_{i.}]. \qquad (\text{~~9~~-33}) \; 11\text{-}36$$

Because

$$E\left[\sum_{t=1}^{T}(\varepsilon_{it} - \bar{\varepsilon}_{i.})^2\right] = (T-1)\sigma_\varepsilon^2,$$

if $\boldsymbol{\beta}$ were observed, then an unbiased estimator of σ_ε^2 based on T observations in group i would be

$$\hat{\sigma}_\varepsilon^2(i) = \frac{\sum_{t=1}^{T}(\varepsilon_{it} - \bar{\varepsilon}_{i.})^2}{T-1}. \qquad (\text{~~9~~-34}) \; 11\text{-}37$$

11-37

Because $\boldsymbol{\beta}$ must be estimated—(9-33) implies that the LSDV estimator is consistent, indeed, unbiased in general—we make the degrees of freedom correction and use the LSDV residuals in

$$s_e^2(i) = \frac{\sum_{t=1}^{T}(e_{it} - \overline{e}_{i.})^2}{T - K - 1}. \qquad (9\text{-}35)$$

(Note that based on the LSDV estimates, $\overline{e}_{i.}$ is actually zero. We will carry it through nonetheless to maintain the analogy to (9-34) where $\overline{\varepsilon}_{i.}$ is not zero but is an estimator of $E[\varepsilon_{it}] = 0$.) We have n such estimators, so we average them to obtain

$$\bar{s}_e^2 = \frac{1}{n}\sum_{i=1}^{n} s_e^2(i) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum_{t=1}^{T}(e_{it} - \overline{e}_{i.})^2}{T - K - 1}\right] = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(e_{it} - \overline{e}_{i.})^2}{nT - nK - n}. \qquad (9\text{-}36)$$

The degrees of freedom correction in $\bar{s}_e^2$ is excessive because it assumes that α and $\boldsymbol{\beta}$ are reestimated for each i. The estimated parameters are the n means $\overline{y}_{i.}$ and the K slopes. Therefore, we propose the unbiased estimator[14]

$$\hat{\sigma}_\varepsilon^2 = s_{LSDV}^2 = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(e_{it} - \overline{e}_{i.})^2}{nT - n - K}. \qquad (9\text{-}37)$$

This is the variance estimator in the fixed effects model in (9-17), appropriately corrected for degrees of freedom. It remains to estimate σ_u^2. Return to the original model specification in (9-32). In spite of the correlation across observations, this is a classical regression model in which the ordinary least squares slopes and variance estimators are both consistent and, in most cases, unbiased. Therefore, using the ordinary least squares residuals from the model with only a single overall constant, we have

$$\text{plim } s_{Pooled}^2 = \text{plim } \frac{\mathbf{e}'\mathbf{e}}{nT - K - 1} = \sigma_\varepsilon^2 + \sigma_u^2. \qquad (9\text{-}38)$$

This provides the two estimators needed for the variance components; the second would be $\hat{\sigma}_u^2 = s_{Pooled}^2 - s_{LSDV}^2$. A possible complication is that this second estimator could be negative. But, recall that for feasible generalized least squares, we do not need an unbiased estimator of the variance, only a consistent one. As such, we may drop the degrees of freedom corrections in (9-37) and (9-38). If so, then the two variance estimators must be nonnegative, since the sum of squares in the LSDV model cannot be larger than that in the simple regression with only one constant term. Alternative estimators have been proposed, all based on this principle of using two different sums of squared residuals.[15] This is a point on which modern software varies greatly. Generally, programs begin with (9-37) and (9-38) to estimate the variance components. What they do next when the estimate of σ_u^2 is nonpositive is far from uniform. Dropping the degrees of freedom correction is a frequently used strategy, but at least one widely used program simply sets σ_u^2 to zero, and others resort to different strategies based on, for example, the group means estimator. The unfortunate implication for the unwary is that different programs can systematically produce different results using the same model and the

AU: Confirm subs & LSDV Pooled are ok italic

[14] A formal proof of this proposition may be found in Maddala (1971) or in Judge et al. (1985, p. 551).

[15] See, for example, Wallace and Hussain (1969), Maddala (1971), Fuller and Battese (1974), and Amemiya (1971).

same data. The practitioner is strongly advised to consult the program documentation for resolution.

There is a remaining complication. If there are any regressors that do not vary within the groups, the LSDV estimator cannot be computed. For example, in a model of family income or labor supply, one of the regressors might be a dummy variable for location, family structure, or living arrangement. Any of these could be perfectly collinear with the fixed effect for that family, which would prevent computation of the LSDV estimator. In this case, it is still possible to estimate the random effects variance components. Let $[\mathbf{b}, a]$ be any consistent estimator of $[\boldsymbol{\beta}, \alpha]$ in (9-32), such as the ordinary least squares estimator. Then, (9-38) provides a consistent estimator of $m_{ee} = \sigma_\varepsilon^2 + \sigma_u^2$. The mean squared residuals using a regression based only on the n group means in (9-32) provides a consistent estimator of $m_{**} = \sigma_u^2 + (\sigma_\varepsilon^2/T)$, so we can use

$$\hat{\sigma}_\varepsilon^2 = \frac{T}{T-1}(m_{ee} - m_{**})$$

$$\hat{\sigma}_u^2 = \frac{T}{T-1}m_{**} - \frac{1}{T-1}m_{ee} = \omega m_{**} + (1-\omega)m_{ee},$$

where $\omega > 1$. As before, this estimator can produce a negative estimate of σ_u^2 that, once again, calls the specification of the model into question. [Note, finally, that the residuals in (9-37) and (9-38) could be based on the same coefficient vector.]

There is, perhaps surprisingly, a simpler way out of the dilemma posed by time-invariant regressors. In (9-33), we find that the group mean deviations estimator still provides a consistent estimator of σ_ε^2. The time-invariant variables fall out of the model so it is not possible to estimate the full coefficient vector $\boldsymbol{\beta}$. But, recall, estimation of $\boldsymbol{\beta}$ is not the objective at this step, estimation of σ_ε^2 is. Therefore, it follows that the residuals from the group mean deviations (LSDV) estimator can still be used to estimate σ_ε^2. By the same logic, the first differences could also be used. (See Section 9.3.5.) The residual variance in the first difference regression would estimate $2\sigma_\varepsilon^2$. These outcomes are irrespective of whether there are time-invariant regressors in the model.

9.5.3 TESTING FOR RANDOM EFFECTS

Breusch and Pagan (1980) have devised a **Lagrange multiplier test** for the random effects model based on the OLS residuals.[16] For

$$H_0: \sigma_u^2 = 0 \quad (\text{or } \mathrm{Corr}[\eta_{it}, \eta_{is}] = 0),$$
$$H_1: \sigma_u^2 \neq 0,$$

the test statistic is

$$\mathrm{LM} = \frac{nT}{2(T-1)}\left[\frac{\sum_{i=1}^n\left[\sum_{t=1}^T e_{it}\right]^2}{\sum_{i=1}^n\sum_{t=1}^T e_{it}^2} - 1\right]^2 = \frac{nT}{2(T-1)}\left[\frac{\sum_{i=1}^n (T\bar{e}_{i.})^2}{\sum_{i=1}^n\sum_{t=1}^T e_{it}^2} - 1\right]^2. \quad \textbf{(9-39)}$$

[16] We have focused thus far strictly on generalized least squares and moments based consistent estimation of the variance components. The LM test is based on maximum likelihood estimation, instead. See Maddala (1971) and Balestra and Nerlove (1966, 2003) for this approach to estimation.

11-39

Under the null hypothesis, the limiting distribution of LM is chi-squared with one degree of freedom.

Not KT "random effects model" not in chap. list

Example ~~9.5~~ 11.6 Testing for Random Effects

We are interested in comparing the random and fixed effects estimators in the Cornwell and Rupert wage equation. As we saw earlier, there are three time-invariant variables in the equation: *Ed*, *Fem*, and *Blk*. As such, we cannot directly compare the two estimators. The **random effects model** can provide separate estimates of the parameters on the time-invariant variables while the fixed effects estimator cannot. For purposes of the illustration, then, we will for the present time confine attention to the restricted common effects model,

KT

$$\ln Wage_{it} = \beta_1 Exp_{it} + \beta_2 Exp_{it}^2 + \beta_3 Wks_{it} + \beta_4 Occ_{it} + \beta_5 Ind_{it} + \beta_6 South_{it} + \beta_7 SMSA_{it} + \beta_8 MS_{it} + \beta_9 Union_{it} + c_i + \varepsilon_{it}.$$

The fixed and random effects models differ in the treatment of c_i.

Least squares estimates of the parameters including a constant term appear in Table ~~9.5~~ 11.6. We then computed the group mean residuals for the seven observations for each individual. The sum of squares of the means is 53.824384. The total sum of squared residuals for the regression is 607.1265. With T and n equal to 7 and 595, respectively, ~~(9-39)~~ 11-42 produces a chi-squared statistic of 3881.34. This far exceeds the 95 percent critical value for the chi-squared distribution with one degree of freedom, 3.84. At this point, we conclude that the classical regression model with a single constant term is inappropriate for these data. The result of the test is to reject the null hypothesis in favor of the random effects model. But, it is best to reserve judgment on that, because there is another competing specification that might induce these same results, the fixed effects model. We will examine this possibility in the subsequent examples.

TB 11.6

TABLE ~~9.5~~ 11.6 Estimates of the Wage Equation

	Pooled Least Squares		*Fixed Effects LSDV*		*Random Effects FGLS*		
Variable	*Estimate*	*Std.Error*[a]	*Estimate*	*Std.Error*	*Estimate*	*Std.Error*	Robust
Exp	0.0361	0.004533	0.1132	0.002471	0.08906	0.002280	0.01276
Exp^2	−0.0006550	0.0001016	−0.0004184	0.0000546	−0.0007577	0.00005036	0.00031
Wks	0.004461	0.001728	0.0008359	0.0005997	0.001066	0.0005939	0.00331
Occ	−0.3176	0.02726	−0.02148	0.01378	−0.1067	0.01269	0.05424
Ind	0.03213	0.02526	0.01921	0.01545	−0.01637	0.01391	0.05303
South	−0.1137	0.02868	−0.001861	0.03430	−0.06899	0.02354	0.05984
SMSA	0.1586	0.02602	−0.04247	0.01943	−0.01530	0.01649	0.05421
MS	0.3203	0.03494	−0.02973	0.01898	−0.02398	0.01711	0.06989
Union	0.06975	0.02667	0.03278	0.01492	0.03597	0.01367	0.05653
Constant	5.8802	0.09673			5.3455	0.04361	0.19866
			Mundlak: Group Means		*Mundlak: Time Varying*		
Exp			−0.08574	0.005821	0.1132	0.002474	
Exp^2			−0.0001168	0.0001281	−0.0004184	0.00005467	
Wks			0.008020	0.004006	0.0008359	0.0006004	
Occ			−0.3321	0.03363	−0.02148	0.01380	
Ind			0.02677	0.03203	0.01921	0.01547	
South			−0.1064	0.04444	−0.001861	0.03434	
SMSA			0.2239	0.03421	0.04247	0.01945	
MS			0.4134	0.03984	−0.02972	0.01901	
Union			0.05637	0.03549	0.03278	0.01494	
Constant					5.7222	0.1906	

[a]Robust standard errors

11-40

With the variance estimators in hand, FGLS can be used to estimate the parameters of the model. All of our earlier results for FGLS estimators apply here. In particular, all that is needed for efficient estimation of the model parameters are consistent estimators of the variance components, and there are several. [See Hsiao (2003), Baltagi (2005), Nerlove (2002), Berzeg (1979), and Maddala and Mount (1973).]

Example ~~9.6~~ 11.7 Estimates of the Random Effects Model

In the previous example, we found the total sum of squares for the least squares estimator was 607.1265. The fixed effects (LSDV) estimates for this model appear in Table ~~9.4~~ 11.5 (and ~~9.5~~ 11.6), where the sum of squares given is 82.26732. Therefore, the moment estimators of the parameters are

$$\hat{\sigma}_\varepsilon^2 + \hat{\sigma}_u^2 = \frac{607.1265}{4165 - 10} = 0.1461195.$$

and

$$\hat{\sigma}_\varepsilon^2 = \frac{82.26732}{4165 - 595 - 9} = 0.0231023.$$

The implied estimator of σ_u^2 is 0.12301719. (No problem of negative variance components has emerged.) The estimate of θ for FGLS is

$$\hat{\theta} = 1 - \sqrt{\frac{0.0231023}{0.0231023 + 7(0.12301719)}} = 0.8383608.$$

FGLS estimates are computed by regressing the partial differences of ln $Wage_{it}$ on the partial differences of the constant and the 9 regressors, using this estimate of θ in (~~9-30~~ 11). Estimates of the parameters using the OLS, fixed effects and random effects estimators appear in Table ~~9.5~~. 11.6.

nine

FN 17

None of the desirable properties of the estimators in the random effects model rely on T going to infinity.[17] Indeed, T is likely to be quite small. The estimator of σ_ε^2 is equal to an average of n estimators, each based on the T observations for unit i. [See (~~9-36~~ 11-39).] Each component in this average is, in principle, consistent. That is, its variance is of order $1/T$ or smaller. Because T is small, this variance may be relatively large. But, each term provides some information about the parameter. The average over the n cross-sectional units has a variance of order $1/(nT)$, which will go to zero if n increases, even if we regard T as fixed. The conclusion to draw is that nothing in this treatment relies on T growing large. Although it can be shown that some consistency results will follow for T increasing, the typical panel data set is based on data sets for which it does not make sense to assume that T increases without bound or, in some cases, at all.[18] As a general proposition, it is necessary to take some care in devising estimators whose properties hinge on whether T is large or not. The widely used conventional ones we have discussed here do not, but we have not exhausted the possibilities.

FN 18

The random effects model was developed by Balestra and Nerlove (1966). Their formulation included a time-specific component, κ_t, as well as the individual effect:

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

[17] See Nickell (1981).

[18] In this connection, Chamberlain (1984) provided some innovative treatments of panel data that, in fact, take T as given in the model and that base consistency results solely on n increasing. Some additional results for dynamic models are given by Bhargava and Sargan (1983).

The extended formulation is rather complicated analytically. In Balestra and Nerlove's study, it was made even more so by the presence of a lagged dependent variable. A full set of results for this extended model, including a method for handling the lagged dependent variable, has been developed.[19] We will turn to this in Section 9.8.

9.5.4 HAUSMAN'S SPECIFICATION TEST FOR THE RANDOM EFFECTS MODEL

At various points, we have made the distinction between fixed and random effects models. An inevitable question is, Which should be used? From a purely practical standpoint, the dummy variable approach is costly in terms of degrees of freedom lost. On the other hand, the fixed effects approach has one considerable virtue. There is little justification for treating the individual effects as uncorrelated with the other regressors, as is assumed in the random effects model. The random effects treatment, therefore, may suffer from the inconsistency due to this correlation between the included variables and the random effect.[20]

The **specification test** devised by Hausman (1978)[21] is used to test for orthogonality of the common effects and the regressors. The test is based on the idea that under the hypothesis of no correlation, both OLS in the LSDV model and GLS are consistent, but OLS is inefficient,[22] whereas under the alternative, OLS is consistent, but GLS is not. Therefore, under the null hypothesis, the two estimates should not differ systematically, and a test can be based on the difference. The other essential ingredient for the test is the covariance matrix of the difference vector, $[\mathbf{b} - \hat{\boldsymbol{\beta}}]$:

$$\text{Var}[\mathbf{b} - \hat{\boldsymbol{\beta}}] = \text{Var}[\mathbf{b}] + \text{Var}[\hat{\boldsymbol{\beta}}] - \text{Cov}[\mathbf{b}, \hat{\boldsymbol{\beta}}] - \text{Cov}[\hat{\boldsymbol{\beta}}, \mathbf{b}]. \tag{9-40}$$

Hausman's essential result is that *the covariance of an efficient estimator with its difference from an inefficient estimator is zero,* which implies that

$$\text{Cov}[(\mathbf{b} - \hat{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}] = \text{Cov}[\mathbf{b}, \hat{\boldsymbol{\beta}}] - \text{Var}[\hat{\boldsymbol{\beta}}] = \mathbf{0}$$

or that

$$\text{Cov}[\mathbf{b}, \hat{\boldsymbol{\beta}}] = \text{Var}[\hat{\boldsymbol{\beta}}].$$

Inserting this result in (9-40) produces the required covariance matrix for the test,

$$\text{Var}[\mathbf{b} - \hat{\boldsymbol{\beta}}] = \text{Var}[\mathbf{b}] - \text{Var}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\Psi}.$$

The chi-squared test is based on the Wald criterion:

$$W = \chi^2[K-1] = [\mathbf{b} - \hat{\boldsymbol{\beta}}]'\hat{\boldsymbol{\Psi}}^{-1}[\mathbf{b} - \hat{\boldsymbol{\beta}}]. \tag{9-41}$$

For $\hat{\boldsymbol{\Psi}}$, we use the estimated covariance matrices of the slope estimator in the LSDV model and the estimated covariance matrix in the random effects model, excluding the

[19] See Balestra and Nerlove (1966), Fomby, Hill, and Johnson (1984), Judge et al. (1985), Hsiao (1986), Anderson and Hsiao (1982), Nerlove (1971a, 2002), and Baltagi (2005).

[20] See Hausman and Taylor (1981) and Chamberlain (1978).

[21] Related results are given by Baltagi (1986).

[22] Referring to the GLS matrix weighted average given earlier, we see that the efficient weight uses θ, whereas OLS sets $\theta = 1$.

11-42

constant term. Under the null hypothesis, W has a limiting chi-squared distribution with $K-1$ degrees of freedom.

The **Hausman test** is a useful device for determining the preferred specification of the common effects model. As developed here, it has one practical shortcoming. The construction in (9-40) conforms to the theory of the test. However, it does not guarantee that the difference of the two covariance matrices will be positive definite in a finite sample. The implication is that nothing prevents the statistic from being negative when it is computed according to (9-41). One can, in that event, conclude that the random effects model is not rejected, since the similarity of the covariance matrices is what is causing the problem, and under the alternative (fixed effects) hypothesis, they would be significantly different. There are, however, several alternative methods of computing the statistic for the Hausman test, some asymptotically equivalent and others actually numerically identical. Baltagi (2005, pp. 65–73) provides an extensive analysis. One particularly convenient form of the test finesses the practical problem noted here. An asymptotically equivalent test statistic is given by

AU: Italics ok in (11-45) and Exm. 11.8?

$$H' = (\hat{\beta}_{LSDV} - \hat{\beta}_{MEANS})' \left[Asy.Var[\hat{\beta}_{LSDV}] + Asy.Var[\hat{\beta}_{MEANS}]\right]^{-1} (\hat{\beta}_{LSDV} - \hat{\beta}_{MEANS}) \quad (9\text{-}42)$$

11-45

where $\hat{\beta}_{MEANS}$ is the group means estimator discussed in Section 9.3.4. As noted, this is one of several equivalent forms of the test. The advantage of this form is that the covariance matrix will always be nonnegative definite.

Example 9.7 Hausman Test for Fixed versus Random Effects

11.8

Using the results of the preceding example, we retrieved the coefficient vector and estimated asymptotic covariance matrix, $\mathbf{b}_{FE}$ and $\mathbf{V}_{FE}$ from the fixed effects results and the first nine elements of $\hat{\beta}_{RE}$ and $\mathbf{V}_{RE}$ (excluding the constant term). The test statistic is

$$H = (\mathbf{b}_{FE} - \hat{\beta}_{RE})'[\mathbf{V}_{FE} - \mathbf{V}_{RE}]^{-1}(\mathbf{b}_{FE} - \hat{\beta}_{RE})$$

The value of the test statistic is 2,636.08. The critical value from the chi-squared table is 4.07, so the null hypothesis of the random effects model is rejected. We conclude that the fixed effects model is the preferred specification for these data. This is an unfortunate turn of events, as the main object of the study is the impact of education, which is a time-invariant variable in this sample. Using (9-42) instead, we obtain a test statistic of 3,177.58. Of course, this does not change the conclusion.

16.919

9.5.5 EXTENDING THE UNOBSERVED EFFECTS MODEL: MUNDLAK'S APPROACH

Even with the Hausman test available, choosing between the fixed and random effects specifications presents a bit of a dilemma. Both specifications have unattractive shortcomings. The fixed effects approach is robust to correlation between the omitted heterogeneity and the regressors, but it proliferates parameters and cannot accommodate time-invariant regressors. The random effects model hinges on an unlikely assumption, that the omitted heterogeneity is uncorrelated with the regressors.

Several authors have suggested modifications of the random effects model that would at least partly overcome its deficit. The failure of the random effects approach is that the mean independence assumption, $E[c_i \mid \mathbf{X}_i] = 0$, is untenable. **Mundlak's** (1978)

11-43

44 45

Imbens and Wooldridge (2007) have argued that in spite of the practical considerations about the Hausman test in (11-41) and (11-42), the test should be based on robust covariance matrices that do not depend on the assumption of the null hypothesis (the random effects model). (I.e., "It makes no sense to report a fully robust variance matrix for FE and RE but then to compute a Hausman test that maintains the full set of RE assumptions.") Their suggested approach amounts to the variable addition test described in the next section, with a robust covariance matrix.

11.5.~~5~~6 EXTENDING THE UNOBSERVED EFFECTS MODEL: MUNDLAK'S APPROACH

Even with the Hausman test available, choosing between the fixed and random effects specifications presents a bit of a dilemma. Both specifications have unattractive shortcomings. The fixed effects approach is robust to correlation between the omitted heterogeneity and the regressors, but it proliferates parameters and cannot accommodate time-invariant regressors. The random effects model hinges on an unlikely assumption, that the omitted heterogeneity is uncorrelated with the regressors. Several authors have suggested modifications of the random effects model that would at least partly overcome its deficit. The failure of the random effects approach is that the mean independence assumption, $E[c_i \mid \mathbf{X}_i] = 0$, is untenable. **Mundlak's** (1978)

KT

11-44

KT

approach would suggest the specification

FN 23

$$E[c_i \mid \mathbf{X}_i] = \bar{\mathbf{x}}_{i.}'\gamma.^{23}$$

Substituting this in the random effects model, we obtain

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\beta + c_i + \varepsilon_{it} \\ &= \mathbf{x}_{it}'\beta + \bar{\mathbf{x}}_{i.}'\gamma + \varepsilon_{it} + (c_i - E[c_i \mid \mathbf{X}_i]) \\ &= \mathbf{x}_{it}'\beta + \bar{\mathbf{x}}_{i.}'\gamma + \varepsilon_{it} + u_i. \end{aligned} \qquad (\cancel{9}\text{-}\cancel{43})\ 11\text{-}46$$

This preserves the specification of the random effects model, but (one hopes) deals directly with the problem of correlation of the effects and the regressors. Note that the additional terms in $\bar{\mathbf{x}}_{i.}'\gamma$ will only include the time-varying variables—the time-invariant variables are already group means. This additional set of estimates is shown in the lower panel of Table ~~9.5 in~~ Example ~~9.6.~~ 11.6.

11.6 Mundlak's approach is frequently used as a compromise between the fixed and random effects models. One side benefit of the specification is that it provides another convenient approach to the Hausman test. As the model is formulated above, the difference between the "fixed effects" model and the "random effects" model is the nonzero γ. As such, a statistical test of the null hypothesis that γ equals zero should provide an alternative approach to the two methods suggested earlier.

Example ~~9.8~~ 11.9 Variable Addition Test for Fixed versus Random Effects

Using the results in Example ~~9.6~~ 11.7, we recovered the subvector of the estimates in the lower half of Table ~~9.5~~ 11.6 corresponding to γ, and the corresponding submatrix of the full covariance matrix. The test statistic is

$$H' = \hat{\gamma}'[\text{Est. Asy. Var}(\hat{\gamma})]^{-1}\hat{\gamma}.$$

The value of the test statistic is ~~297.17~~ 3193.69. The critical value from the chi-squared table for nine degrees of freedom is ~~14.07~~ 16.919 so the null hypothesis of the random effects model is rejected. We conclude as before that the fixed effects estimator is the preferred specification for this model.

~~9.6~~ 11 NONSPHERICAL DISTURBANCES AND ROBUST COVARIANCE ESTIMATION

Because the models considered here are extensions of the classical regression model, we can treat heteroscedasticity in the same way that we did in Chapter ~~8~~ 9. That is, we can compute the ordinary or feasible generalized least squares estimators and obtain an appropriate robust covariance matrix estimator, or we can impose some structure on the disturbance variances and use generalized least squares. In the panel data settings,

for example,

[23] Other analyses, ~~e.g.,~~ Chamberlain (1982) and Wooldridge (2002a), interpret the linear function as the *projection* of c_i on the group means, rather than the conditional mean. The difference is that we need not make any particular assumptions about the conditional mean function while there always exists a linear projection. The conditional mean interpretation does impose an additional assumption on the model, but brings considerable simplification. Several authors have analyzed the extension of the model to projection on the full set of individual observations rather than the means. The additional generality provides the bases of several other estimators including minimum distance [Chamberlain (1982)], GMM [Arellano and Bover (1995)], and constrained seemingly unrelated regressions and three-stage least squares [Wooldridge (2002a)].

AU: OK to spell out "e.g." ?

11-45

(1983) treatment of a form 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.

Example 10.5 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. The central equation of the model is

$$\ln Q_{it,j} = \beta_0 + \beta_1 \ln P\_natural\ gas_{it,j} + \beta_2 \ln P\_electricity_{it,j} + \beta_3 \ln P\_fuel\ oil_{it,j} + \beta_4 \ln per\ capita\ income_{it,j} + \beta_5 \ln Q_{i,t-1,j} + w_{it,j}$$

$$w_{it,j} = \varepsilon_{it,j} + u_{i,j} + v_{t,j}$$

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.

11.5.7 ~~10.3.2~~ EXTENDING THE RANDOM AND FIXED EFFECTS MODELS: CHAMBERLAIN'S APPROACH

The linear unobserved effects model is

$$y_{it} = c_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}. \quad \text{(11-47) } \cancel{(10\text{-}22)}$$

AU: Term "random effects" already in chap. here also?

The **random effects** model assumes that $E[c_i \mid \mathbf{X}_i] = \alpha$, 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 $E[\mathbf{w}_i\mathbf{w}_i' \mid \mathbf{X}_i] = \sigma_\varepsilon^2\mathbf{I}_T + \sigma_u^2\mathbf{ii}'$, 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 11.5 ~~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 $E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' \mid \mathbf{X}_i] = \sigma_\varepsilon^2\mathbf{I}_T$ assumes that the correlation across periods is equal for all pairs of observations, and arises solely through the persistent c_i. In Section 10.2.6 ~~10.2.8~~, we estimated the equivalent model with an unrestricted covariance matrix, $E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' \mid \mathbf{X}_i] = \boldsymbol{\Sigma}$. The implication is that the random effects treatment includes two restrictive assumptions, mean independence, $E[c_i \mid \mathbf{X}_i] = \alpha$, and homoscedasticity, $E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' \mid \mathbf{X}_i] = \sigma_\varepsilon^2\mathbf{I}_T$. [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 $E[\mathbf{w}_i\mathbf{w}_i' \mid \mathbf{X}_i] = \sigma_\varepsilon^2\mathbf{I}_T + \sigma_u^2\mathbf{ii}'$.

11-46

Au: Two terms already KTs in chap. Here also?

The mean independence assumption is the major shortcoming of the **random effects model.** The central feature of the fixed effects model in Section ~~9~~11.4 is the possibility that $E[c_i \mid \mathbf{X}_i]$ is a nonconstant $g(\mathbf{X}_i)$. As such, least squares regression of y_{it} on $\mathbf{x}_{it}$ produces an inconsistent estimator of $\boldsymbol{\beta}$. The dummy variable model considered in Section ~~9.4~~ 11 is the natural alternative. The **fixed effects** approach has the advantage of dispensing with the unlikely assumption that c_i and $\mathbf{x}_{it}$ are uncorrelated. However, it has the shortcoming of requiring estimation of the n "parameters," α_i.

Au: Term "projections" not in chap. list.

Chamberlain (1982, 1984) and Mundlak (1978) suggested alternative approaches that lie between these two. Their modifications of the fixed effects model augment it with the **projections** of c_i on all the rows of $\mathbf{X}_i$ (Chamberlain) or the group means (Mundlak). (See Sections ~~4.2.2~~ 11 and ~~9~~11.5.5.) Consider the first of these, and assume (as it requires) a **balanced panel** of T observations per group. For purposes of this development, we will assume $T = 3$. The generalization will be obvious at the conclusion. Then, the projection suggested by Chamberlain is

Au: Term "balanced panel" already a KT in this chap. Here also?

$$c_i = \alpha + \mathbf{x}_{i1}'\boldsymbol{\gamma}_1 + \mathbf{x}_{i2}'\boldsymbol{\gamma}_2 + \mathbf{x}_{i3}'\boldsymbol{\gamma}_3 + r_i \qquad (\cancel{10\text{-}23})\ 11\text{-}45$$

where now, by construction, r_i is orthogonal to $\mathbf{x}_{it}$.[17] Insert (~~10-23~~ 11-44) into (~~10-22~~ 11-45) to obtain

$$y_{it} = \alpha + \mathbf{x}_{i1}'\boldsymbol{\gamma}_1 + \mathbf{x}_{i2}'\boldsymbol{\gamma}_2 + \mathbf{x}_{i3}'\boldsymbol{\gamma}_3 + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} + r_i.$$

Estimation of the $1 + 3K + K$ parameters of this model presents a number of complications. [We do note, this approach has the potential to (wildly) proliferate parameters. For our quite small regional productivity model in Example ~~10.2~~ 11.19, the original model with six main coefficients plus the treatment of the constants becomes a model with $1 + 6 + 17(6) = 109$ parameters to be estimated.]

If only the n observations for period 1 are used, then the parameter vector,

$$\boldsymbol{\theta}_1 = \alpha, (\boldsymbol{\beta} + \boldsymbol{\gamma}_1), \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3 = \alpha, \boldsymbol{\pi}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3, \qquad (\cancel{10\text{-}24})\ 11\text{-}46$$

can be estimated consistently, albeit inefficiently, by ordinary least squares. The "model" is

$$y_{i1} = \mathbf{z}_{i1}'\boldsymbol{\theta}_1 + w_{i1}, i = 1, \ldots, n.$$

Collecting the n observations, we have

$$\mathbf{y}_1 = \mathbf{Z}_1\boldsymbol{\theta}_1 + \mathbf{w}_1.$$

If, instead, only the n observations from period 2 or period 3 are used, then OLS estimates, in turn,

$$\boldsymbol{\theta}_2 = \alpha, \boldsymbol{\gamma}_1, (\boldsymbol{\beta} + \boldsymbol{\gamma}_2), \boldsymbol{\gamma}_3 = \alpha, \boldsymbol{\gamma}_1, \boldsymbol{\pi}_2, \boldsymbol{\gamma}_3,$$

or

$$\boldsymbol{\theta}_3 = \alpha, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, (\boldsymbol{\beta} + \boldsymbol{\gamma}_3) = \alpha, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\pi}_3.$$

[17] There are some fine points here that can only be resolved theoretically. If the projection in (~~10-23~~ 11-45) is not the conditional mean, then we have $E[r_i \times \mathbf{x}_{it}] = 0, t = 1, \ldots, T$ but not $E[r_i \mid \mathbf{X}_i] = 0$. This does not affect the asymptotic properties of the FGLS estimator to be developed here, although it does have implications, e.g., for unbiasedness. Consistency will hold regardless. The assumptions behind (~~10-23~~ 11-45) do not include that $\text{Var}[r_i \mid \mathbf{X}_i]$ is homoscedastic. It might not be. This *could* be investigated empirically. The implication here concerns efficiency, not consistency. The FGLS estimator to be developed here would remain consistent, but a GMM estimator would be more efficient—see Chapter ~~15~~. Moreover, without homoscedasticity, it is not certain that the FGLS estimator suggested here is more efficient than OLS (with a robust covariance matrix estimator). Our intent is to begin the investigation here. Further details can be found in Chamberlain (1984) and, e.g., Im, Ahn, Schmidt, and Wooldridge (1999).

for example, (for example,)

Au: OK to spell out "e.g." – 2X?

11-47

It remains to reconcile the multiple estimates of the same parameter vectors. In terms of the preceding layouts above, we have the following:

OLS Estimates:	$a_1, \mathbf{p}_1, \mathbf{c}_{2,1}, \mathbf{c}_{3,1},$	$a_2\, \mathbf{c}_{1,2}, \mathbf{p}_2, \mathbf{c}_{3,2},$	$a_3, \mathbf{c}_{1,3}, \mathbf{c}_{2,3}, \mathbf{p}_3;$
Estimated Parameters:	$\alpha, (\beta+\gamma_1), \gamma_2, \gamma_3,$	$\alpha, \gamma_1, (\beta+\gamma_2), \gamma_3,$	$\alpha, \gamma_1, \gamma_2, (\beta+\gamma_3);$
Structural Parameters:	$\alpha, \beta, \gamma_1, \gamma_2, \gamma_3.$		

50 ~~(10-25)~~ 11-~~52~~ 47

Chamberlain suggested a minimum distance estimator (MDE). For this problem, the MDE is essentially a weighted average of the several estimators of each part of the parameter vector. We will examine the MDE for this application in more detail in Chapter ~~15~~ 13. (For another simpler application of minimum distance estimation that shows the "weighting" procedure at work, see the reconciliation of four competing estimators of a single parameter at the end of Example ~~9.15~~ 11.20.) There is an alternative way to formulate the estimator that is a bit more transparent. For the first period,

$$\mathbf{y}_1 = \begin{pmatrix} y_{1,1} \\ y_{2,1} \\ \vdots \\ y_{n,1} \end{pmatrix} = \begin{bmatrix} 1 & \mathbf{x}_{1,1} & \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \mathbf{x}_{1,3} \\ 1 & \mathbf{x}_{2,2} & \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \mathbf{x}_{2,3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \mathbf{x}_{n,1} & \mathbf{x}_{n,1} & \mathbf{x}_{n,1} & \mathbf{x}_{n,1} \end{bmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} + \begin{pmatrix} r_{1,1} \\ r_{2,1} \\ \vdots \\ r_{n,1} \end{pmatrix} = \tilde{\mathbf{X}}_1\theta + \mathbf{r}_1. \qquad (\cancel{10\text{-}26})\ 11\text{-}\cancel{53}\,48$$

51

We treat this as the first equation in a T equation seemingly unrelated regressions model. The second equation, for period 2, is the same (same coefficients), with the data from the second period appearing in the blocks, then likewise for period 3 (and periods $4, \ldots, T$ in the general case). Stacking the data for the T equations (periods), we have

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_T \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{X}}_1 \\ \tilde{\mathbf{X}}_2 \\ \vdots \\ \tilde{\mathbf{X}}_T \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma_1 \\ \vdots \\ \gamma_T \end{pmatrix} + \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_T \end{pmatrix} = \tilde{\mathbf{X}}\theta + \mathbf{r}, \qquad (\cancel{10\text{-}27})\ 11\text{-}\cancel{54}\,49$$

52

where $E[\tilde{\mathbf{X}}'\mathbf{r}] = \mathbf{0}$ and (by assumption), $E[\mathbf{rr}' \mid \tilde{\mathbf{X}}] = \boldsymbol{\Sigma} \otimes \mathbf{I}_n$. With the homoscedasticity assumption for $r_{i,t}$, this is precisely the application in Section 10.2.~~4~~6. The parameters can be estimated by FGLS as shown in Section 10.2.~~3~~6

Example ~~10.6~~ 11.10 Hospital Costs

Carey (1997) examined hospital costs for a sample of 1,733 hospitals observed in five years, 1987–1991. The model estimated is

$$\begin{aligned} \ln(TC/P)_{it} &= \alpha_i + \beta_D DIS_{it} + \beta_O OPV_{it} + \beta_3 ALS_{it} + \beta_4 CM_{it} \\ &+ \beta_5 DIS_{it}^2 + \beta_6 DIS_{it}^3 + \beta_7 OPV_{it}^2 + \beta_8 OPV_{it}^3 \\ &+ \beta_9 ALS_{it}^2 + \beta_{10} ALS_{it}^3 + \beta_{11} DIS_{it} \times OPV_{it} \\ &+ \beta_{12} FA_{it} + \beta_{13} HI_{it} + \beta_{14} HT_i + \beta_{15} LT_i + \beta_{16} Large_i \\ &+ \beta_{17} Small_i + \beta_{18} NonProfit_i + \beta_{19} Profit_i \\ &+ \varepsilon_{it}, \end{aligned}$$

where

TC	= total cost,
P	= input price index,
DIS	= discharges,
OPV	= outpatient visits,
ALS	= average length of stay,
CM	= case mix index,
FA	= fixed assets,
HI	= Hirfindahl index of market concentration at county level,
HT	= dummy for high teaching load hospital,
LT	= dummy variable for low teaching load hospital,
Large	= dummy variable for large urban area,
Small	= dummy variable for small urban area,
Nonprofit	= dummy variable for nonprofit hospital,
Profit	= dummy variable for for profit hospital.

We have used subscripts "D" and "O" for the coefficients on DIS and OPV as these will be isolated in the following discussion. The model employed in the study is that in (~~10-22~~ 11-47) and (~~10-23~~ 11-45). Initial OLS estimates are obtained for the full cost function in each year. SUR estimates are then obtained using a restricted version of the Chamberlain system. This second step involved a hybrid model that modified (~~10-24~~ 11-46) so that in each period the coefficient vector was

$$\theta_t = [\alpha_t, \beta_{Dt}(\gamma), \beta_{Ot}(\gamma), \beta_{3t}(\gamma), \beta_{4t}(\gamma), \beta_{5t}, \ldots, \beta_{19t}]$$

where $\beta_{Dt}(\gamma)$ indicates that all five years of the variable (DIS_{it}) are included in the equation and, likewise for $\beta_{Ot}(\gamma)(OPV)$, $\beta_{3t}(\gamma)(ALS)$ and $\beta_{4t}(\gamma)(CM)$. This is equivalent to using

$$c_i = \alpha + \Sigma_{t=1987}^{1991}(DIS, OPV, ALS, CM)'_{it}\gamma_t + r_i$$

in (~~10-23~~ 11-45).

The unrestricted SUR system estimated at the second step provides multiple estimates of the various model parameters. For example, each of the five equations provides an estimate of $(\beta_5, \ldots, \beta_{19})$. The author added one more layer to the model in allowing the coefficients on DIS_{it} and OPV_{it} to vary over time. Therefore, the structural parameters of interest are $(\beta_{D1}, \ldots, \beta_{D5})$, $(\gamma_{D1} \ldots, \gamma_{D5})$ (the coefficients on DIS) and $(\beta_{O1}, \ldots, \beta_{O5})$, $(\gamma_{O1} \ldots, \gamma_{O5})$ (the coefficients on OPV). There are, altogether, 20 parameters of interest. The SUR estimates produce, in each year (equation), parameters on DIS for the five years and on OPV for the five years, so there are a total of 50 estimates. Reconciling all of them means imposing a total of 30 restrictions. Table ~~10.2~~ 11.7 shows the relationships for the time varying parameter on DIS_{it} in the five-equation model. The numerical values reported by the author are shown following the theoretical results. A similar table would apply for the coefficients on OPV, ALS, and CM. (In the latter two, the β coefficient was not assumed to be time varying.) It can be seen in the table, for example, that there are directly four different estimates of $\gamma_{D,87}$ in the second to fifth equations, and likewise for each of the other parameters. Combining the entries in Table ~~10.2~~ 11.7 with the counterpart for the coefficients on OPV, we see 50 SUR/FGLS estimates to be used to estimate 20 underlying parameters. The author used a minimum distance approach to reconcile the different estimates. We will return to this example in ~~Chapter 17~~ Example 13.6, where we will develop the MDE in more detail.

TB 11.7