11-49

11.7

TABLE ~~10.2~~ Coefficient Estimates in SUR Model for Hospital Costs

Equation	DIS87	DIS88	DIS89	DIS90	DIS91
	Coefficient on Variable in the Equation				
SUR87	$\beta_{D,87}+\gamma_{D,87}$ 1.76	$\gamma_{D,88}$ 0.116	$\gamma_{D,89}$ −0.0881	$\gamma_{D,90}$ 0.0570	$\gamma_{D,91}$ −0.0617
SUR88	$\gamma_{D,87}$ 0.254	$\beta_{D,88}+\gamma_{D,88}$ 1.61	$\gamma_{D,89}$ −0.0934	$\gamma_{D,90}$ 0.0610	$\gamma_{D,91}$ −0.0514
SUR89	$\gamma_{D,87}$ 0.217	$\gamma_{D,88}$ 0.0846	$\beta_{D,89}+\gamma_{D,89}$ 1.51	$\gamma_{D,90}$ 0.0454	$\gamma_{D,91}$ −0.0253
SUR90	$\gamma_{D,87}$ 0.179	$\gamma_{D,88}$ 0.0822[a]	$\gamma_{D,89}$ 0.0295	$\beta_{D,90}+\gamma_{D,90}$ 1.57	$\gamma_{D,91}$ 0.0244
SUR91	$\gamma_{D,87}$ 0.153	$\gamma_{D,88}$ 0.0363	$\gamma_{D,89}$ −0.0422	$\gamma_{D,90}$ 0.0813	$\beta_{D,91}+\gamma_{D,91}$ 1.70

[a]The value reported in the published paper is 8.22. The correct value is 0.0822. (Personal communication from the author.)

10.4 SYSTEMS OF DEMAND EQUATIONS: SINGULAR SYSTEMS

Most of the recent applications of the multivariate regression model[18] have been in the context of **systems of demand equations,** either commodity demands or factor demands in studies of production.

Example 10.7 Stone's Expenditure System

Stone's expenditure system[19] based on a set of logarithmic commodity demand equations, income Y, and commodity prices p_n is

$$\log q_i = \alpha_i + \eta_i \log\left(\frac{Y}{P}\right) + \sum_{j=1}^{M} \eta_{ij}^* \log\left(\frac{p_j}{P}\right),$$

where P is a generalized (share-weighted) price index, η_i is an income elasticity, and η_{ij}^* is a compensated price elasticity. We can interpret this system as the demand equation in real expenditure and real prices. The resulting set of equations constitutes an econometric model in the form of a set of seemingly unrelated regressions. In estimation, we must account for a number of restrictions including homogeneity of degree one in income, $\Sigma_i S_i \eta_i = 1$, and symmetry of the matrix of compensated price elasticities, $\eta_{ij}^* = \eta_{ji}^*$, where S_i is the budget share for good i.

Other examples include the system of factor demands and factor cost shares from production, which we shall consider again later. In principle, each is merely a particular application of the model of the Section 10.2. But some special problems arise in these settings. First, the parameters of the systems are generally constrained across equations. That is, the unconstrained model is inconsistent with the underlying

~~[18]Note the distinction between the multivariate or multiple-equation model discussed here and the multiple regression model.~~

[19]A very readable survey of the estimation of systems of commodity demands is Deaton and Muellbauer (1980). The example discussed here is taken from their Chapter 3 and the references to Stone's (1954a,b) work cited therein. Deaton (1986) is another useful survey. A counterpart for production function modeling is Chambers (1988). More recent developments in the specification of systems of demand equations include Chavez and Segerson (1987), Brown and Walker (1995), and Fry, Fry, and McLaren (1996).

11-50

approach would suggest the specification

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

Substituting this in the random effects model, we obtain

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

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.

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 Variable Addition Test for Fixed versus Random Effects

Using the results in Example 9.6, we recovered the subvector of the estimates in the lower half of Table 9.5 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. The critical value from the chi-squared table for nine degrees of freedom is 14.07, 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.

11 ~~9~~.6 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~~. 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, 9

[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)].

there is greater flexibility for the second of these without making strong assumptions about the nature of the heteroscedasticity.

9.6.1 ROBUST ESTIMATION OF THE FIXED EFFECTS MODEL

As noted in Section 8.3.2, in a panel data set, the correlation across observations within a group is likely to be a more substantial influence on the estimated covariance matrix of the least squares estimator than is heteroscedasticity. This is evident in the estimates in Table 9.1. In the fixed (or random) effects model, the intent of explicitly including the common effect in the model is to account for the source of this correlation. However, accounting for the common effect in the model does not remove heteroscedasticity—it centers the conditional mean properly. Here, we consider the straightforward extension of White's estimator to the fixed and random effects models.

In the fixed effects model, the full regressor matrix is $\mathbf{Z} = [\mathbf{X}, \mathbf{D}]$. The White heteroscedasticity consistent covariance matrix for OLS—that is, for the fixed effects estimator—is the lower right block of the partitioned matrix

$$\text{Est. Asy. Var}[\mathbf{b}, \mathbf{a}] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{E}^2\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1},$$

where $\mathbf{E}$ is a diagonal matrix of least squares (fixed effects estimator) residuals. This computation promises to be formidable, but fortunately, it works out very simply. The White estimator for the slopes is obtained just by using the data in group mean deviation form [see (9-14) and (9-17)] in the familiar computation of $\mathbf{S}_0$ [see (8-26) and (8-27)]. Also, the disturbance variance estimator in (9-17) is the counterpart to the one in (8-20), which we showed that after the appropriate scaling of $\boldsymbol{\Omega}$ was a consistent estimator of $\sigma^2 = \text{plim}[1/(nT)]\sum_{i=1}^{n}\sum_{t=1}^{T}\sigma_{it}^2$. The implication is that we may still use (9-17) to estimate the variances of the fixed effects.

A somewhat less general but useful simplification of this result can be obtained if we assume that the disturbance variance is constant within the ith group. If $E[\varepsilon_{it}^2 \,|\, \mathbf{Z}_i] = \sigma_i^2$, then, with a panel of data, σ_i^2 is estimable by $\mathbf{e}_i'\mathbf{e}_i/T$ using the least squares residuals. The center matrix in Est. Asy. Var[$\mathbf{b}, \mathbf{a}$] may be replaced with $\sum_i(\mathbf{e}_i'\mathbf{e}_i/T)\mathbf{Z}_i'\mathbf{Z}_i$. Whether this estimator is preferable is unclear. If the groupwise model is correct, then it and the White estimator will estimate the same matrix. On the other hand, if the disturbance variances do vary within the groups, then this revised computation may be inappropriate.

Arellano (1987) and Arellano and Bover (1995) have taken this analysis a step further. If one takes the ith group as a whole, then we can treat the observations in

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \alpha_i\mathbf{i}_T + \boldsymbol{\varepsilon}_i$$

as a generalized regression model with disturbance covariance matrix $\boldsymbol{\Omega}_i$. We saw in Section 8.3.2 that a model this general, with no structure on $\boldsymbol{\Omega}$, offered little hope for estimation, robust or otherwise. But the problem is more manageable with a panel data set where correlation across units can be assumed to be zero. As before, let $\mathbf{X}_{i*}$ denote the data in group mean deviation form. The counterpart to $\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}$ here is

$$\mathbf{X}_*'\boldsymbol{\Omega}\mathbf{X}_* = \sum_{i=1}^{n}(\mathbf{X}_{i*}'\boldsymbol{\Omega}_i\mathbf{X}_{i*}).$$

By the same reasoning that we used to construct the White estimator in Chapter 8, we can consider estimating $\boldsymbol{\Omega}_i$ with the sample of one, $\mathbf{e}_i\mathbf{e}_i'$. As before, it is not consistent estimation of the individual Ω_is that is at issue, but estimation of the sum. If n is large enough, then we could argue that

$$\begin{aligned}
\operatorname{plim} \frac{1}{nT}\mathbf{X}_*'\Omega\mathbf{X}_* &= \operatorname{plim} \frac{1}{nT}\sum_{i=1}^{n}\mathbf{X}_{i*}'\Omega_i\mathbf{X}_{*i} \\
&= \operatorname{plim} \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\mathbf{X}_{*i}'\mathbf{e}_i\mathbf{e}_i'\mathbf{X}_{*i} \\
&= \operatorname{plim} \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}e_{it}e_{is}\mathbf{x}_{*it}\mathbf{x}_{*is}'\right).
\end{aligned}$$

11-53
11-50
~~(9-44)~~

This is the extension of (~~9~~ 11-3) to the fixed effects case.

~~9~~ 11.6.2 HETEROSCEDASTICITY IN THE RANDOM EFFECTS MODEL

Because the random effects model is a generalized regression model with a known structure, OLS with a robust estimator of the asymptotic covariance matrix is not the best use of the data. The GLS estimator is efficient whereas the OLS estimator is not. If a perfectly general covariance structure is assumed, then one might simply use Arellano's estimator described in the preceding section with a single overall constant term rather than a set of fixed effects. But, within the setting of the random effects model, $\eta_{it} = \varepsilon_{it} + u_i$, allowing the disturbance variance to vary across groups would seem to be a useful extension.

A series of papers, notably Mazodier and Trognon (1978), Baltagi and Griffin (1988), and the recent monograph by Baltagi (2005, pp. 77–79) suggest how one might allow the group-specific component u_i to be heteroscedastic. But, empirically, there is an insurmountable problem with this approach. In the final analysis, all estimators of the variance components must be based on sums of squared residuals, and, in particular, an estimator of σ_{ui}^2 would be estimated using a set of residuals from the distribution of u_i. However, the data contain only a single observation on u_i repeated in each observation in group i. So, the estimators presented, for example, in Baltagi (2001), use, in effect, one residual in each case to estimate σ_{ui}^2. What appears to be a mean squared residual is only $(1/T)\sum_{t=1}^{T}\hat{u}_i^2 = \hat{u}_i^2$. The properties of this estimator are ambiguous, but efficiency seems unlikely. The estimators do not converge to any population figure as the sample size, even T, increases. [The counterpoint is made in Hsiao (2003, p. 56).] Heteroscedasticity in the unique component, ε_{it} represents a more tractable modeling possibility.

In Section ~~9.5.1~~ 11.5.2, we introduced heteroscedasticity into estimation of the random effects model by allowing the group sizes to vary. But the estimator there (and its feasible counterpart in the next section) would be the same if, instead of $\theta_i = 1 - \sigma_\varepsilon/(T_i\sigma_u^2 + \sigma_\varepsilon^2)^{1/2}$, we were faced with

$$\theta_i = 1 - \frac{\sigma_{\varepsilon i}}{\sqrt{\sigma_{\varepsilon i}^2 + T_i\sigma_u^2}}.$$

Therefore, for computing the appropriate feasible generalized least squares estimator, once again we need only devise consistent estimators for the variance components and

11-53

then apply the GLS transformation shown earlier. One possible way to proceed is as follows: Because pooled OLS is still consistent, OLS provides a usable set of residuals. Using the OLS residuals for the specific groups, we would have, for each group,

$$\widehat{\sigma_{\varepsilon i}^2 + u_i^2} = \frac{\mathbf{e}_i'\mathbf{e}_i}{T}.$$

The residuals from the dummy variable model are purged of the individual specific effect, u_i, so $\sigma_{\varepsilon i}^2$ may be consistently (in T) estimated with

$$\widehat{\sigma_{\varepsilon i}^2} = \frac{\mathbf{e}_i^{lsdv\prime}\mathbf{e}_i^{lsdv}}{T}$$

where $e_{it}^{lsdv} = y_{it} - \mathbf{x}_{it}'\mathbf{b}^{lsdv} - a_i$. Combining terms, then,

$$\hat{\sigma}_u^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{\mathbf{e}_i^{ols\prime}\mathbf{e}_i^{ols}}{T}\right) - \left(\frac{\mathbf{e}_i^{lsdv\prime}\mathbf{e}_i^{lsdv}}{T}\right)\right] = \frac{1}{n}\sum_{i=1}^{n}\widehat{(u_i^2)}.$$

We can now compute the FGLS estimator as before.

11 ~~9~~.6.3 AUTOCORRELATION IN PANEL DATA MODELS

Section 20.10

Serial correlation of regression disturbances will be considered in detail in ~~Chapter 19~~ 20. 20
Rather than defer the topic in connection to panel data to Chapter ~~19~~ 20, we will briefly note it here. As we saw in Section ~~9~~ 11.3.2 and Example ~~9~~ 11.1, "autocorrelation"—that is, correlation across the observations in the groups in a panel—is likely to be a substantive feature of the model. Our treatment of the effect there, however, was meant to accommodate autocorrelation in its broadest sense, that is, nonzero covariances across observations in a group. The results there would apply equally to clustered observations, as observed in Section ~~9~~ 11.3.3. An important element of that specification was that with clustered data, there might be no obvious structure to the autocorrelation. When the panel data set consists explicitly of groups of time series, and especially if the time series are relatively long as in Example ~~9.9~~ 11.11, one might want to begin to invoke the more detailed, structured time series models which are discussed in Chapter ~~19~~ 20.

9.7 EXTENSIONS OF THE RANDOM EFFECTS MODEL

In spite of its strong assumption of mean independence of the individual effects and the regressors, the random effects model, perhaps augmented with Mundlak's extension (Section 9.5.5), provides the preferred framework for much of the empirical literature. (As we will explore in several applications later in the book, the fixed effects model has a significant shortcoming of its own.) This section will describe a few common extensions of the model. The nested random effects model was suggested earlier as an approach for analyzing hierarchical data sets. We will describe some of the received estimation techniques and applications for nested effects. Spatial autocorrelation is a natural application of panel models in which the correlation across observations relates to their distance from each other in space rather than time. Finally, we will take a brief look at dynamic panel data models and nonstationary panels.

11.6.4 CLUSTER (AND PANEL) ROBUST COVARIANCE MATRICES FOR FIXED AND RANDOM EFFECTS ESTIMATORS

AU: Confirm italic subs in (11-54)–(11-57)

As suggested earlier, in situations in which cluster corrections are appropriate, there might be a residual correlation within groups ~~not~~ that is not fully accounted for by a generalized least squares estimator or a fixed effects model. A counterpart to (11-4) for the fixed and random effects estimators is straightforward to construct based on results we have already obtained.

For the fixed effects estimator, based on (11-13) (14) and (11-19) (20), we have

$$\mathbf{b}_{LSDV} = \left[\sum_{g=1}^{G}\sum_{i=1}^{n_g}(\Delta(1)\mathbf{x}_{ig})(\Delta(1)\mathbf{x}_{ig})'\right]^{-1}\left[\sum_{g=1}^{G}\sum_{i=1}^{n_g}(\Delta(1)\mathbf{x}_{ig})(\Delta(1)y_{ig})\right] \quad (11\text{-}51)\ 54$$

where $\Delta(1)\mathbf{x}_{it} = \mathbf{x}_{it} - (1)\overline{\mathbf{x}}_i$ is the deviation of $\mathbf{x}_{it}$ from one times the group mean vector. The motivation for the "(1)" will be evident shortly. In the same fashion as (11-3), we will construct a robust covariance matrix estimator using

$$\begin{aligned} Est.Asy.Var[\mathbf{b}_{LSDV}] = {} & \left[\sum_{g=1}^{G}\sum_{i=1}^{n_g}(\Delta(1)\mathbf{x}_{ig})(\Delta(1)\mathbf{x}_{ig})'\right]^{-1} \times \\ & \left[\sum_{g=1}^{G}\left\{\sum_{i=1}^{n_g}(\Delta(1)\mathbf{x}_{ig})e_{ig}\right\}\left\{\sum_{i=1}^{n_g}(\Delta(1)\mathbf{x}_{ig})e_{ig}\right\}'\right] \times \\ & \left[\sum_{g=1}^{G}\sum_{i=1}^{n_g}(\Delta(1)\mathbf{x}_{ig})(\Delta(1)\mathbf{x}_{ig})'\right]^{-1}. \end{aligned} \quad (11\text{-}52)\ 55$$

This estimator is equivalent to (11-3) based on the data in deviations from their cluster means. (With a slight change in notation, it becomes a robust estimator for the covariance matrix of the fixed effects estimator.) From (11-29) (32) and (11-30) (33), the GLS estimator of β for the random effects model is

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{GLS} & = \left[\sum_{g=1}^{G}\mathbf{X}_g'\boldsymbol{\Sigma}_g^{-1}\mathbf{X}_g\right]^{-1}\left[\sum_{g=1}^{G}\mathbf{X}_g'\boldsymbol{\Sigma}_g^{-1}\mathbf{y}_g\right] \\ & = \left[\sum_{g=1}^{G}\sum_{i=1}^{n_g}(\Delta(\theta_g)\mathbf{x}_{ig})(\Delta(\theta_g)\mathbf{x}_{ig})'\right]^{-1}\left[\sum_{g=1}^{G}\sum_{i=1}^{n_g}(\Delta(\theta_g)\mathbf{x}_{ig})(\Delta(\theta_g)y_{ig})\right], \end{aligned} \quad (11\text{-}53)\ 56$$

where $\theta_g = 1 - \left(\sigma_\varepsilon / \sqrt{\sigma_\varepsilon^2 + n_g\sigma_u^2}\right)$. It follows that the estimator of the asymptotic covariance matrix would be

$$\begin{aligned} Est.Asy.Var[\hat{\boldsymbol{\beta}}_{GLS}] = {} & \left[\sum_{g=1}^{G}\sum_{i=1}^{n_g}(\Delta(\theta_g)\mathbf{x}_{ig})(\Delta(\theta_g)\mathbf{x}_{ig})'\right]^{-1} \times \\ & \left[\sum_{g=1}^{G}\left\{\sum_{i=1}^{n_g}(\Delta(\theta_g)\mathbf{x}_{ig})e_{ig}\right\}\left\{\sum_{i=1}^{n_g}(\Delta(\theta_g)\mathbf{x}_{ig})e_{ig}\right\}'\right] \times \\ & \left[\sum_{g=1}^{G}\sum_{i=1}^{n_g}(\Delta(\theta_g)\mathbf{x}_{ig})(\Delta(\theta_g)\mathbf{x}_{ig})'\right]^{-1}. \end{aligned} \quad (11\text{-}54)\ 57$$

See, also, Cameron and Trivedi (2005, pp. 838–839).

11-55

AU: Confirm table number is OK

TB 11.8

8

Example 11.11 Robust Standard Errors for Fixed and Random Effects Estimators

Table 11.7 presents the estimates of the fixed random effects models that appear in Tables 11.5 and 11.6. The correction of the standard errors results in a fairly substantial change in the estimates. The effect is especially promounced in the random effects case, where the estimated standard errors increase by a factor of five or more.

8

Table 11.11 Cluster Corrections for Fixed and Random Effects Estimators

Note minus signs

	Fixed Effects			Random Effects		
Variable	Estimate	Std.Error	Robust	Estimate	Std.Error	Robust
Constant				5.3455	0.04361	0.19866
Exp	0.1132	0.002471	0.00437	0.08906	0.002280	0.01276
Exp^2	-0.00042	0.000055	0.000089	-0.0007577	0.00005036	0.00031
Wks	0.00084	0.000600	0.00094	0.001066	0.0005939	0.00331
Occ	-0.02148	0.01378	0.02052	-0.1067	0.01269	0.05424
Ind	0.01921	0.01545	0.02450	-0.01637	0.01391	0.053003
South	-0.00186	0.03430	0.09646	-0.06899	0.02354	0.05984
SMSA	-0.04247	0.01942	0.03185	-0.01530	0.01649	0.05421
MS	-0.02973	0.01898	0.02902	-0.02398	0.01711	0.06984
Union	0.03278	0.01492	0.02708	0.03597	0.01367	0.05653

11.7 SPATIAL AUTOCORRELATION

The nested random effects structure in Example 11.12 was motivated by an expectation that effects of neighboring states would spill over into each other, creating a sort of correlation across space, rather than across time as we have focused on thus far. The effect should be common in cross-region studies, such as in agriculture, urban economics, and regional science. Recent studies of the phenomenon include Case's (1991) study of expenditure patterns, Bell and Bockstael's (2000) study of real estate prices, and Baltagi and Li's (2001) analysis of R&D spillovers. Models of **spatial autocorrelation** [see Anselin (1988, 2001) for the canonical reference and LeSage and Pace (2009) for a recent survey], are constructed to formalize this notion.

A model with spatial autocorrelation can be formulated as follows: The regression model takes the familiar panel structure,

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

The common u_i is the usual unit (e.g., country) effect. The correlation across space is implied by the spatial autocorrelation structure

$$\varepsilon_{it} = \lambda \sum_{j=1}^{n} W_{ij}\varepsilon_{jt} + v_t.$$

The scalar λ is the **spatial autoregression coefficient.** The elements W_{ij} are spatial (or **contiguity**) weights that are assumed known. The elements that appear in the sum above are a row of the spatial weight or **contiguity matrix**, $\mathbf{W}$, so that for the n units, we have

$$\boldsymbol{\varepsilon}_t = \lambda\mathbf{W}\boldsymbol{\varepsilon}_t + \mathbf{v}_t, \mathbf{v}_t = v_t\mathbf{i}.$$

The structure of the model is embodied in the symmetric weight matrix, $\mathbf{W}$. Consider for an example counties or states arranged geographically on a grid or some linear scale such as a line from one coast of the country to another. Typically W_{ij} will equal one for i, j pairs that are neighbors and zero otherwise. Alternatively, W_{ij} may reflect distances across space, so that W_{ij} decreases with increases in $|i - j|$. This would be similar to a temporal autocorrelation matrix. Assuming that $|\lambda|$ is less than one, and that the elements of $\mathbf{W}$ are such that $(\mathbf{I} - \lambda\mathbf{W})$ is nonsingular, we may write

$$\boldsymbol{\varepsilon}_t = (\mathbf{I}_n - \lambda\mathbf{W})^{-1}\mathbf{v}_t,$$

so for the n observations at time t,

$$\mathbf{y}_t = \mathbf{X}_t\boldsymbol{\beta} + (\mathbf{I}_n - \lambda\mathbf{W})^{-1}\mathbf{v}_t + \mathbf{u}.$$

We further assume that u_i and v_i have zero means, variances σ_u^2 and σ_v^2 and are independent across countries and of each other. It follows that a generalized regression model applies to the n observations at time t;

$$E[\mathbf{y}_t \mid \mathbf{X}_t] = \mathbf{X}_t\boldsymbol{\beta},$$
$$\text{Var}[\mathbf{y}_t \mid \mathbf{X}_t] = (\mathbf{I}_n - \lambda\mathbf{W})^{-1}[\sigma_v^2\mathbf{i}\mathbf{i}'](\mathbf{I}_n - \lambda\mathbf{W})^{-1} + \sigma_u^2\mathbf{I}_n.$$

At this point, estimation could proceed along the lines of Chapter ~~8~~ 9, save for the need to estimate λ. There is no natural residual based estimator of λ. Recent treatments of this model have added a normality assumption and employed maximum likelihood

11-57

methods. [The log likelihood function for this model and numerous references appear in Baltagi (2005, p. 196). Extensive analysis of the estimation problem is given in Bell and Bockstael (2000).]

A natural first step in the analysis is a test for spatial effects. The standard procedure for a cross section is Mōran's (1950) I statistic, which would be computed for each set of residuals, $\mathbf{e}_t$, using

58

$$I_t = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}(e_{it} - \bar{e}_t)(e_{jt} - \bar{e}_t)}{\left(\sum_{i=1}^{n} \sum_{j=1}^{n} W_{i,j}\right) \sum_{i=1}^{n} (e_{it} - \bar{e}_t)^2}. \quad (11\text{-}55)$$

For a panel of T independent sets of observations, $\bar{I} = \frac{1}{T} \sum_{t=1}^{T} I_t$ would use the full set of information. A large sample approximation to the variance of the statistic under the null hypothesis of no spatial autocorrelation is

59

$$V^2 = \frac{1}{T} \frac{n^2 \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}^2 + 3\left(\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}\right)^2 - n \sum_{i=1}^{n} \left(\sum_{j=1}^{n} W_{ij}\right)^2}{(n^2 - 1)\left(\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}\right)^2}. \quad (11\text{-}56)$$

The statistic $\bar{I}/V$ will converge to standard normality under the null hypothesis and can form the basis of the test. (The assumption of independence across time is likely to be dubious at best, however.) Baltagi, Song, and Koh (2003) identify a variety of LM tests based on the assumption of normality. Two that apply to cross section analysis [See Bell and Bockstael (2000, p. 78)] are

$$LM(1) = \frac{(\mathbf{e}'\mathbf{W}\mathbf{e}/s^2)^2}{tr(\mathbf{W}'\mathbf{W} + \mathbf{W}^2)}$$

for spatial autocorrelation and

$$LM(2) = \frac{(\mathbf{e}'\mathbf{W}\mathbf{y}/s^2)^2}{\mathbf{b}'\mathbf{X}'\mathbf{W}\mathbf{M}\mathbf{W}\mathbf{X}\mathbf{b}/s^2 + tr(\mathbf{W}'\mathbf{W} + \mathbf{W}^2)}$$

for spatially lagged dependent variables, where $\mathbf{e}$ is the vector of OLS residuals, $s^2 = \mathbf{e}'\mathbf{e}/n$ and $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. [See Anselin and Hudak (1992).]

Anselin (1988) identifies several possible extensions of the spatial model to dynamic regressions. A "pure space-recursive model" specifies that the autocorrelation pertains to neighbors in the previous period:

$$y_{it} = \gamma[\mathbf{W}\mathbf{y}_{t-1}]_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}.$$

A "time-space recursive model" specifies dependence that is purely autoregressive with respect to neighbors in the previous period:

$$y_{it} = \rho y_{i,t-1} + \gamma[\mathbf{W}\mathbf{y}_{t-1}]_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}.$$

A "time-space simultaneous" model specifies that the spatial dependence is with respect to neighbors in the current period:

$$y_{it} = \rho y_{i,t-1} + [\lambda \mathbf{W}\mathbf{y}_t]_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}.$$

Finally, a "time-space dynamic model" specifies that autoregression depends on neighbors in both the current and last period:

$$y_{it} = \rho y_{i,t-1} + [\lambda \mathbf{W}\mathbf{y}_t]_i + \gamma[\mathbf{W}\mathbf{y}_{t-1}]_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}. \quad (11.12)$$

Example ~~9.10~~ Spatial Autocorrelation in Real Estate Sales

Bell and Bockstael analyzed the problem of modeling spatial autocorrelation in large samples. This is likely to become an increasingly common problem with GIS (geographic information system) data sets. The central problem is maximization of a likelihood function that involves a sparse matrix, $(\mathbf{I} - \lambda\,\mathbf{W})$. Direct approaches to the problem can encounter severe inaccuracies in evaluation of the inverse and determinant. Kelejian and Prucha (1999) have developed a moment-based estimator for λ that helps to alleviate the problem. Once the estimate of λ is in hand, estimation of the spatial autocorrelation model is done by FGLS. The authors applied the method to analysis of a cross section of 1,000 residential sales in Anne Arundel County, Maryland, from 1993 to 1996. The parcels sold all involved houses built within one year prior to the sale. GIS software was used to measure attributes of interest.

The model is

$$\begin{aligned}
\ln \textit{Price} = \alpha &+ \beta_1 \ln \textit{Assessed value (LIV)} \\
&+ \beta_2 \ln \textit{Lot size (LLT)} \\
&+ \beta_3 \ln \textit{Distance in km to Washington, DC (LDC)} \\
&+ \beta_4 \ln \textit{Distance in km to Baltimore (LBA)} \\
&+ \beta_5 \%\ \textit{land surrounding parcel in publicly owned space (POPN)} \\
&+ \beta_6 \%\ \textit{land surrounding parcel in natural privately owned space (PNAT)} \\
&+ \beta_7 \%\ \textit{land surrounding parcel in intensively developed use (PDEV)} \\
&+ \beta_8 \%\ \textit{land surrounding parcel in low density residential use (PLOW)} \\
&+ \beta_9\ \textit{Public sewer service (1 if existing or planned, 0 if not) (PSEW)} \\
&+ \varepsilon.
\end{aligned}$$

(Land surrounding the parcel is all parcels in the GIS data whose centroids are within 500 meters of the transacted parcel.) For the full model, the specification is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$
$$\boldsymbol{\varepsilon} = \lambda\mathbf{W}\boldsymbol{\varepsilon} + \mathbf{v}.$$

The authors defined four contiguity matrices:

W1: W_{ij} = 1/distance between i and j if distance < 600 meters, 0 otherwise,
W2: W_{ij} = 1 if distance between i and j < 200 meters, 0 otherwise,
W3: W_{ij} = 1 if distance between i and j < 400 meters, 0 otherwise,
W4: W_{ij} = 1 if distance between i and j < 600 meters, 0 othewise.

All contiguity matrices were row-standardized. That is, elements in each row are scaled so that the row sums to one. One of the objectives of the study was to examine the impact of row standardization on the estimation. It is done to improve the numerical stability of the optimization process. Because the estimates depend numerically on the normalization, it is not completely innocent.

TB 11.9

Test statistics for spatial autocorrelation based on the OLS residuals are shown in Table ~~9.7~~ 11.9. (These are taken from the authors' Table 3.) The Moran statistics are distributed as standard normal while the LM statistics are distributed as chi-squared with one degree of freedom. All but the LM(2) statistic for W3 are larger than the 99% critical value from the respective table, so we would conclude that there is evidence of spatial autocorrelation. Estimates from some of the regressions are shown in Table ~~9.8~~ 11.10. In the remaining results in the study, the

TB 11.10

TABLE ~~9.7~~ Test Statistics for Spatial Autocorrelation

	W1	*W2*	*W3*	*W4*
Moran's *I*	7.89	9.67	13.66	6.88
LM(1)	49.95	84.93	156.48	36.46
LM(2)	7.40	17.22	2.33	7.42

11.10

TABLE ~~9.8~~ Estimated Spatial Regression Models

	OLS		*FGLS*[a]		*Spatial based on W1 ML*		*Spatial based on W1 Gen. Moments*	
Parameter	*Estimate*	*Std.Err.*	*Estimate*	*Std.Err.*	*Estimate*	*Std.Err.*	*Estimate*	*Std.Err.*
α	4.7332	0.2047	4.7380	0.2048	5.1277	0.2204	5.0648	0.2169
β_1	0.6926	0.0124	0.6924	0.0214	0.6537	0.0135	0.6638	0.0132
β_2	0.0079	0.0052	0.0078	0.0052	0.0002	0.0052	0.0020	0.0053
β_3	−0.1494	0.0195	−0.1501	0.0195	−0.1774	0.0245	−0.1691	0.0230
β_4	−0.0453	0.0114	−0.0455	0.0114	−0.0169	0.0156	−0.0278	0.0143
β_5	−0.0493	0.0408	−0.0484	0.0408	−0.0149	0.0414	−0.0269	0.0413
β_6	0.0799	0.0177	0.0800	0.0177	0.0586	0.0213	0.0644	0.0204
β_7	0.0677	0.0180	0.0680	0.0180	0.0253	0.0221	0.0394	0.0211
β_8	−0.0166	0.0194	−0.0168	0.0194	−0.0374	0.0224	−0.0313	0.0215
β_9	−0.1187	0.0173	−0.1192	0.0174	−0.0828	0.0180	−0.0939	0.0179
λ	—	—	—	—	0.4582	0.0454	0.3517	—

[a]The author reports using a heteroscedasticity model $\sigma_i^2 \times f(LIV_i, LIV_i^2)$. The function $f(.)$ is not identified.

authors find that the outcomes are somewhat sensitive to the specification of the spatial weight matrix, but not particularly so to the method of estimating λ.

11.13 ***Example ~~9.11~~ Spatial Lags in Health Expenditures***

Moscone, Knapp, and Tosetti (2007) investigated the determinants of mental health expenditure over six years in 148 British local authorities using two forms of the spatial correlation model to incorporate possible interaction among authorities as well as unobserved spatial heterogeneity. The models estimated, in addition to pooled regression and a random effects model, were as follows. The first is a model with **spatial lags:** (KT)

$$\mathbf{y}_t = \gamma_t \mathbf{i} + \rho \mathbf{W} \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta} + \mathbf{u} + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{u}$ is a 148×1 vector of random effects and $\mathbf{i}$ is a 148×1 column of ones. For each local authority,

$$y_{it} = \gamma_t + \rho(\mathbf{w}_i' \mathbf{y}_t) + \mathbf{x}_{it}' \boldsymbol{\beta} + u_i + \varepsilon_{it},$$

where $\mathbf{w}_i'$ is the ith row of the contiguity matrix, $\mathbf{W}$. Contiguities were defined in $\mathbf{W}$ as one if the locality shared a border or vertex and zero otherwise. (The authors also experimented with other contiguity matrices based on "sociodemographic" differences.) The second model estimated is of **spatial error correlation** (KT)

$$\mathbf{y}_t = \gamma_t \mathbf{i} + \mathbf{X}_t \boldsymbol{\beta} + \mathbf{u} + \boldsymbol{\varepsilon}_t,$$
$$\boldsymbol{\varepsilon}_t = \lambda \mathbf{W} \boldsymbol{\varepsilon}_t + \mathbf{v}_t.$$

Avi KT "spatial error correlation" not in chap. list

11-60

For each local authority, this model implies

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

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

9

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

9.8 PARAMETER HETEROGENEITY

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

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

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

TABLE 12.2 Nonlinear Least Squares and Instrumental Variable Estimates

	Instrumental Variables		Least Squares	
Parameter	*Estimate*	*Standard Error*	*Estimate*	*Standard Error*
α	627.031	26.6063	468.215	22.788
β	0.040291	0.006050	0.0971598	0.01064
γ	1.34738	0.016816	1.24892	0.1220
σ	57.1681	—	49.87998	—
$\mathbf{e}'\mathbf{e}$	650,369.805	—	495,114.490	—

Example 12.6 Instrumental Variables Estimates of the Consumption Function

The consumption function in Section 11.3.1 was estimated by nonlinear least squares without accounting for the nature of the data that would certainly induce correlation between $\mathbf{X}^0$ and ε. As we did earlier, we will reestimate this model using the technique of instrumental variables. For this application, we will use the one-period lagged value of consumption and one- and two-period lagged values of income as instrumental variables estimates. Table 12.2 reports the nonlinear least squares and instrumental variables estimates. Because we are using two periods of lagged values, two observations are lost. Thus, the least squares estimates are not the same as those reported earlier.

The instrumental variable estimates differ considerably from the least squares estimates. The differences can be deceiving, however. Recall that the MPC in the model is $\beta\gamma Y^{\gamma-1}$. The 2000.4 value for *DPI* that we examined earlier was 6634.9. At this value, the instrumental variables and least squares estimates of the MPC are 1.1543 with an estimated standard error of 0.01234 and 1.08406 with an estimated standard error of 0.008694, respectively. These values do differ a bit but less than the quite large differences in the parameters might have led one to expect. We do note that the IV estimate is considerably greater than the estimate in the linear model, 0.9217 (and greater than one, which seems a bit implausible).

11.8 ENDOGENEITY ~~AND THE HAUSMAN AND TAYLOR~~ ~~12.8 PANEL DATA APPLICATIONS~~ ~~INSTRUMENTAL VARIABLE~~ ESTIMATOR

AU: Term "panel data" has already been a KT in this chap. KT again?

Recent **panel data** applications have relied heavily on the methods of instrumental variables ~~that we are developing here~~. We will develop this methodology in detail in Chapter ~~15~~ 13 where we consider generalized method of moments (GMM) estimation. At this point, we can examine two major building blocks in this set of methods, Hausman and Taylor's (1981) estimator for the random effects model and Bhargava and Sargan's (1983) proposals for estimating a dynamic panel data model. These two tools play a significant role in the GMM estimators of dynamic panel models in Chapter ~~15~~ 13.

~~12.8.1 INSTRUMENTAL VARIABLES ESTIMATION OF THE RANDOM EFFECTS MODEL—THE HAUSMAN AND TAYLOR ESTIMATOR~~ 11.8.1 Hausman and Taylor's Instrumental Variables Estimator

Recall the original specification of the linear model for panel data in (~~9~~ 11-1):

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\alpha} + \varepsilon_{it}. \qquad \text{(~~12-27~~ 11-57 60)}$$

The random effects model is based on the assumption that the unobserved person-specific effects, $\mathbf{z}_i$, are uncorrelated with the included variables, $\mathbf{x}_{it}$. This assumption is a major shortcoming of the model. However, the random effects treatment does allow the model to contain observed time invariant characteristics, such as demographic

11-62

AU: This term is not in chap. list

characteristics, while the fixed effects model does not—if present, they are simply absorbed into the fixed effects. **Hausman and Taylor's** (1981) **estimator** for the random effects model suggests a way to overcome the first of these while accommodating the second.

Their model is of the form:

$$y_{it} = \mathbf{x}_{1it}'\boldsymbol{\beta}_1 + \mathbf{x}_{2it}'\boldsymbol{\beta}_2 + \mathbf{z}_{1i}'\boldsymbol{\alpha}_1 + \mathbf{z}_{2i}'\boldsymbol{\alpha}_2 + \varepsilon_{it} + u_i$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2')'$. In this formulation, all individual effects denoted $\mathbf{z}_i$ are observed. As before, unobserved individual effects that are contained in $\mathbf{z}_i'\boldsymbol{\alpha}$ in (~~12-27~~) are contained in the person specific random term, u_i. Hausman and Taylor define four sets of *observed* variables in the model:

11-57 60

$\mathbf{x}_{1it}$ is K_1 variables that are time varying and uncorrelated with u_i,
$\mathbf{z}_{1i}$ is L_1 variables that are time invariant and uncorrelated with u_i,
$\mathbf{x}_{2it}$ is K_2 variables that are time varying and are correlated with u_i,
$\mathbf{z}_{2i}$ is L_2 variables that are time invariant and are correlated with u_i.

The assumptions about the random terms in the model are

$$\begin{aligned}
&E[u_i \mid \mathbf{x}_{1it}, \mathbf{z}_{1i}] = 0 \text{ though } E[u_i \mid \mathbf{x}_{2it}, \mathbf{z}_{2i}] \neq 0,\\
&\text{Var}[u_i \mid \mathbf{x}_{1it}, \mathbf{z}_{1i}, \mathbf{x}_{2it}, \mathbf{z}_{2i}] = \sigma_u^2,\\
&\text{Cov}[\varepsilon_{it}, u_i \mid \mathbf{x}_{1it}, \mathbf{z}_{1i}, \mathbf{x}_{2it}, \mathbf{z}_{2i}] = 0,\\
&\text{Var}[\varepsilon_{it} + u_i \mid \mathbf{x}_{1it}, \mathbf{z}_{1i}, \mathbf{x}_{2it}, \mathbf{z}_{2i}] = \sigma^2 = \sigma_\varepsilon^2 + \sigma_u^2,\\
&\text{Corr}[\varepsilon_{it} + u_i, \varepsilon_{is} + u_i \mid \mathbf{x}_{1it}, \mathbf{z}_{1i}, \mathbf{x}_{2it}, \mathbf{z}_{2i}] = \rho = \sigma_u^2/\sigma^2.
\end{aligned}$$

Note the crucial assumption that one can distinguish sets of variables $\mathbf{x}_1$ and $\mathbf{z}_1$ that are uncorrelated with u_i from $\mathbf{x}_2$ and $\mathbf{z}_2$ which are not. The likely presence of $\mathbf{x}_2$ and $\mathbf{z}_2$ is what complicates specification and estimation of the random effects model in the first place.

insert (A) on next page 11-63 msp

By construction, any OLS or GLS estimators of this model are inconsistent when the model contains variables that are correlated with the random effects. Hausman and Taylor have proposed an instrumental variables estimator that uses only the information within the model (i.e., as already stated). The strategy for estimation is based on the following logic: First, by taking deviations from group means, we find that

$$y_{it} - \bar{y}_{i.} = (\mathbf{x}_{1it} - \bar{\mathbf{x}}_{1i.})'\boldsymbol{\beta}_1 + (\mathbf{x}_{2it} - \bar{\mathbf{x}}_{2i.})'\boldsymbol{\beta}_2 + \varepsilon_{it} - \bar{\varepsilon}_{i.}, \qquad \text{11-58 61 } \cancel{(12\text{-}28)}$$

both parts of

which implies that $\boldsymbol{\beta}$ can be consistently estimated by least squares, *in spite of the correlation between* $\mathbf{x}_2$ *and u.* This is the familiar, fixed effects, least squares dummy variable estimator—the transformation to deviations from group means removes from the model the part of the disturbance that is correlated with $\mathbf{x}_{2it}$. Now, in the original model, Hausman and Taylor show that the group mean deviations can be used as $(K_1 + K_2)$ instrumental variables for estimation of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$. That is the implication of (~~12-28~~). Because $\mathbf{z}_1$ is uncorrelated with the disturbances, it can likewise serve as a set of L_1 instrumental variables. That leaves a necessity for L_2 instrumental variables. The authors show that the group means for $\mathbf{x}_1$ can serve as these remaining instruments, and the model will be identified so long as K_1 is greater than or equal to L_2. *For identification purposes, then,* K_1 *must be at least as large as* L_2. As usual, **feasible GLS** is better than OLS, and available. Likewise, FGLS is an improvement over simple instrumental variable estimation of the model, which is consistent but inefficient.

11-58 61

KT

AU: KT "feasible GLS" is not in chap. list

A

Insert on msp 11-62 where indicated

11-63

umlaut

that

We note in passing, that we can contrast the four assumptions with those made in Plümper and Troeger's (2007) FEVD formulation in Section 11.4.5 which, in the notation of this formulation, would be that $\mathbf{x}_{1it}$ and $\mathbf{x}_{2it}$ are time varying and both freely correlated with u_i while $\mathbf{z}_{1i}$ and $\mathbf{z}_{2i}$ are time invariant and are both uncorrelated with u_i. For both formulations, (11-58) applies. The two approaches differ in the additional moment conditions, $E[variable \times (u_i + \varepsilon_{it})] = 0$, that are used to identify the parameters α_1 and α_2.

(61) end of insert

11-64

The authors propose the following set of steps for consistent and efficient estimation:

Step 1. Obtain the LSDV (fixed effects) estimator of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ based on $\mathbf{x}_1$ and $\mathbf{x}_2$. The residual variance estimator from this step is a consistent estimator of σ_ε^2.

Step 2. Form the within-groups residuals, e_{it}, from the LSDV regression at step 1. Stack the group means of these residuals in a full-sample-length data vector. Thus, $e_{it}^* = \bar{e}_{i.} = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \mathbf{x}_{it}'\mathbf{b}_w)$, $t = 1, \ldots, T$, $i = 1, \ldots, n$. (The individual constant term, a_i, is not included in e_{it}^*.) These group means are used as the dependent variable in an instrumental variable regression on $\mathbf{z}_1$ and $\mathbf{z}_2$ with instrumental variables $\mathbf{z}_1$ and $\mathbf{x}_1$. (Note the identification requirement that K_1, the number of variables in $\mathbf{x}_1$ be at least as large as L_2, the number of variables in $\mathbf{z}_2$.) The time-invariant variables are each repeated T times in the data matrices in this regression. This provides a consistent estimator of $\boldsymbol{\alpha}$.

Step 3. The residual variance in the regression in step 2 is a consistent estimator of $\sigma^{*2} = \sigma_u^2 + \sigma_\varepsilon^2/T$. From this estimator and the estimator of σ_ε^2 in step 1, we deduce an estimator of $\sigma_u^2 = \sigma^{*2} - \sigma_\varepsilon^2/T$. We then form the weight for feasible GLS in this model by forming the estimate of

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

Step 4. The final step is a weighted instrumental variable estimator. Let the full set of variables in the model be

$$\mathbf{w}_{it}' = (\mathbf{x}_{1it}', \mathbf{x}_{2it}', \mathbf{z}_{1i}', \mathbf{z}_{2i}').$$

Collect these nT observations in the rows of data matrix $\mathbf{W}$. The transformed variables for GLS are, as before when we first fit the random effects model,

$$\mathbf{w}_{it}^{*\prime} = \mathbf{w}_{it}' - \hat{\theta}\overline{\mathbf{w}}_{i.}' \quad \text{and} \quad y_{it}^* = y_{it} - \hat{\theta}\bar{y}_{i.}$$

where $\hat{\theta}$ denotes the sample estimate of θ. The transformed data are collected in the rows data matrix $\mathbf{W}^*$ and in column vector $\mathbf{y}^*$. Note in the case of the time-invariant variables in $\mathbf{w}_{it}$, the group mean is the original variable, and the transformation just multiplies the variable by $1 - \hat{\theta}$. The instrumental variables are

$$\mathbf{v}_{it}' = [(\mathbf{x}_{1it} - \bar{\mathbf{x}}_{1i.})', (\mathbf{x}_{2it} - \bar{\mathbf{x}}_{2i.})', \mathbf{z}_{1i}', \bar{\mathbf{x}}_{1i.}'].$$

These are stacked in the rows of the $nT \times (K_1 + K_2 + L_1 + K_1)$ matrix $\mathbf{V}$. Note for the third and fourth sets of instruments, the time-invariant variables and group means are repeated for each member of the group. The instrumental variable estimator would be

$$(\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\alpha}}')_{\text{IV}}' = [(\mathbf{W}^{*\prime}\mathbf{V})(\mathbf{V}'\mathbf{V})^{-1}(\mathbf{V}'\mathbf{W}^*)]^{-1}[(\mathbf{W}^{*\prime}\mathbf{V})(\mathbf{V}'\mathbf{V})^{-1}(\mathbf{V}'\mathbf{y}^*)].^{25} \qquad \text{11-59 } \cancel{(12\text{-}29)}$$

FN 25

25 ~~15~~ Note that the FGLS random effects estimator would be $(\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\alpha}}')_{RE}' = [\mathbf{W}^{*\prime}\mathbf{W}^*]^{-1}\mathbf{W}^{*\prime}\mathbf{y}^*$.

The instrumental variable estimator is consistent if the data are not weighted, that is, if $\mathbf{W}$ rather than $\mathbf{W}^*$ is used in the computation. But, this is inefficient, in the same way that OLS is consistent but inefficient in estimation of the simpler random effects model.

Example ~~12.7~~ 11.16 14 The Returns to Schooling

The economic returns to schooling have been a frequent topic of study by econometricians. The PSID and NLS data sets have provided a rich source of panel data for this effort. In wage (or log wage) equations, it is clear that the economic benefits of schooling are correlated with latent, unmeasured characteristics of the individual such as innate ability, intelligence, drive, or perseverance. As such, there is little question that simple random effects models based on panel data will suffer from the effects noted earlier. The fixed effects model is the obvious alternative, but these rich data sets contain many useful variables, such as race, union membership, and marital status, which are generally time invariant. Worse yet, the variable most of interest, years of schooling, is also time invariant. Hausman and Taylor (1981) proposed the estimator described here as a solution to these problems. The authors studied the effect of schooling on (the log of) wages using a random sample from the PSID of 750 men aged 25–55, observed in two years, 1968 and 1972. The two years were chosen so as to minimize the effect of serial correlation apart from the persistent unmeasured individual effects. The variables used in their model were as follows:

AU: Confirm first 2 variables in list are OK

Experience = age – years of schooling – 5,
Years of schooling,
Bad Health = a dummy variable indicating general health,
Race = a dummy variable indicating nonwhite (70 of 750 observations),
Union = a dummy variable indicating union membership,
Unemployed = a dummy variable indicating previous year's unemployment.

The model also included a constant term and a period indicator. [The coding of the latter is not given, but any two distinct values, including 0 for 1968 and 1 for 1972, would produce identical results. (Why?)]

The primary focus of the study is the coefficient on schooling in the log wage equation. Because schooling and, probably, Experience and Unemployed are correlated with the latent effect, there is likely to be serious bias in conventional estimates of this equation. Table ~~12.3~~ 11.11 reports some of their reported results. The OLS and random effects GLS results in the first two columns provide the benchmark for the rest of the study. The schooling coefficient is estimated at 0.0669, a value which the authors suspected was far too small. As we saw earlier, even in the presence of correlation between measured and latent effects, in this model, the LSDV estimator provides a consistent estimator of the coefficients on the time varying variables. Therefore, we can use it in the **Hausman specification test** for correlation between the included variables and the latent heterogeneity. The calculations are shown in Section ~~9~~ 11.5.4, result (~~9-42~~ 11-45). Because there are three variables remaining in the LSDV equation, the chi-squared statistic has three degrees of freedom. The reported value of 20.2 is far larger than the 95 percent critical value of 7.81, so the results suggest that the random effects model is misspecified.

TB 11.11

AU: This term has already been a KT in this chap. Here also?

Hausman and Taylor proceeded to reestimate the log wage equation using their proposed estimator. The fourth and fifth sets of results in Table ~~12.3~~ 11.11 present the instrumental variable estimates. The specification test given with the fourth set of results suggests that the procedure has produced the desired result. The hypothesis of the modified random effects model is now not rejected; the chi-squared value of 2.24 is much smaller than the critical value. The schooling variable is treated as endogenous (correlated with u_i) in both cases. The difference between the two is the treatment of Unemployed and Experience. In the preferred equation, they are included in $\mathbf{x}_2$ rather than $\mathbf{x}_1$. The end result of the exercise is, again, the coefficient on schooling, which has risen from 0.0669 in the worst specification (OLS) to 0.2169 in the last one, a difference of over 200 percent. As the authors note, at the same time, the measured effect of race nearly vanishes.

11-66

11-11

TABLE 12.8 Estimated Log Wage Equations

	Variables	OLS	GLS/RE	LSDV	HT/IV-GLS	HT/IV-GLS
x_1	Experience	0.0132	0.0133	0.0241	0.0217	
		(0.0011)[a]	(0.0017)	(0.0042)	(0.0031)	
	Bad health	−0.0843	−0.0300	−0.0388	−0.0278	−0.0388
		(0.0412)	(0.0363)	(0.0460)	(0.0307)	(0.0348)
	Unemployed Last Year	−0.0015	−0.0402	−0.0560	−0.0559	
		(0.0267)	(0.0207)	(0.0295)	(0.0246)	
	Time	NR[b]	NR	NR	NR	NR
x_2	Experience					0.0241
						(0.0045)
	Unemployed					−0.0560
						(0.0279)
z_1	Race	−0.0853	−0.0878		−0.0278	−0.0175
		(0.0328)	(0.0518)		(0.0752)	(0.0764)
	Union	0.0450	0.0374		0.1227	0.2240
		(0.0191)	(0.0296)		(0.0473)	(0.2863)
	Schooling	0.0669	0.0676			
		(0.0033)	(0.0052)			
	Constant	NR	NR	NR	NR	NR
z_2	Schooling				0.1246	0.2169
					(0.0434)	(0.0979)
	σ_ε	0.321	0.192	0.160	0.190	0.629
	$\rho = \sqrt{\sigma_u^2/(\sigma_u^2+\sigma_\varepsilon^2)}$		0.632		0.661	0.817
	Spec. Test [3]		20.2		2.24	0.00

[a]Estimated asymptotic standard errors are given in parentheses.
[b]NR indicates that the coefficient estimate was not reported in the study.

12.8.2 DYNAMIC PANEL DATA MODELS—THE ANDERSON/HSIAO AND ARELLANO/BOND ESTIMATORS

A leading contemporary application of the methods of this chapter and Chapter 9 is the **dynamic panel data model,** which we now write

$$y_{it} = \mathbf{x}_{it}'\beta + \delta y_{i,t-1} + c_i + \varepsilon_{it}.$$

Several applications are described in Example 9.18. The basic assumptions of the model are

1. Strict exogeneity: $E[\varepsilon_{it} \mid \mathbf{X}_i, c_i] = 0$,
2. Homoscedasticity: $E[\varepsilon_{it}^2 \mid \mathbf{X}_i, c_i] = \sigma_\varepsilon^2$,
3. Nonautocorrelation: $E[\varepsilon_{it}\varepsilon_{is} \mid \mathbf{X}_i, c_i] = 0$ if $t \neq s$,
4. Uncorrelated observations: $E[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}_i, c_i, \mathbf{X}_j, c_j] = 0$ for $i \neq j$ and for all t and s,

where the rows of the $T \times K$ data matrix $\mathbf{X}_i$ are $\mathbf{x}_{it}'$. We will not assume mean independence. The "effects" may be fixed or random, so we allow

$$E[c_i \mid \mathbf{X}_i] = g(\mathbf{X}_i).$$

(See Section 9.2.1.) We will also assume a fixed number of periods, T, for convenience. The treatment here (and in the literature) can be modified to accommodate unbalanced panels, but it is a bit inconvenient. (It involves the placement of zeros at various places

~~A subtle problem arises in obtaining results useful for characterizing the properties of estimators of the model in (9-66). The asymptotic results based on large n and large T are not necessarily obtainable simultaneously, and great care is needed in deriving the asymptotic behavior of useful statistics. Phillips and Moon (1999, 2000) are standard references on the subject.~~

~~We will return to the topic of nonstationary data in Chapter 22. This is an emerging literature, most of which is well beyond the level of this text. We will rely on the several detailed received surveys, such as Bannerjee (1999), Smith (2000), and Baltagi and Kao (2000) to fill in the details.~~

11.8.2 ~~9.9~~ CONSISTENT ESTIMATION OF DYNAMIC PANEL DATA MODELS: ANDERSON AND HSIAO'S IV ESTIMATOR

~~As prelude to the further developments of Chapters 12 and 13,~~ We consider ~~return to~~ a homogeneous dynamic panel data model,

$$y_{it} = \gamma y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + c_i + \varepsilon_{it}, \qquad \text{11-63 } \cancel{(9\text{-}67)}$$

where c_i is, as in the preceding sections of this chapter, individual unmeasured heterogeneity, that may or may not be correlated with $\mathbf{x}_{it}$. We consider methods of estimation for this model when T is fixed and relatively small, and n may be large and increasing.

Pooled OLS is obviously inconsistent. Rewrite (~~9-67~~ 11-63) as

$$y_{it} = \gamma y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + w_{it}. \qquad \cancel{(9\text{-}68)}$$

The disturbance in this pooled regression may be correlated with $\mathbf{x}_{it}$, but either way, it is surely correlated with $y_{i,t-1}$. By substitution,

$$\text{Cov}[y_{i,t-1}, (c_i + \varepsilon_{it})] = \sigma_c^2 + \gamma\text{Cov}[y_{i,t-2}, (c_i + \varepsilon_{it})],$$

and so on. By repeated substitution, it can be seen that for $|\gamma| < 1$ and moderately large T,

$$\text{Cov}[y_{i,t-1}, (c_i + \varepsilon_{it})] \approx \sigma_c^2/(1-\gamma). \qquad \text{11-64 } \cancel{(9\text{-}69)}$$

[It is useful to obtain this result from a different direction. If the stochastic process that is generating (y_{it}, c_i) is *stationary*, then $\text{Cov}[y_{i,t-1}, c_i] = \text{Cov}[y_{i,t-2}, c_i]$, from which we would obtain (~~9-69~~ 11-64) directly. The assumption $|\gamma| < 1$ would be required for stationarity. We will return to this subject in Chapters ~~20 and 21~~ 21 and 22.] Consequently, OLS and GLS are inconsistent. The fixed effects approach does not solve the problem either. Taking deviations from individual means, we have

$$y_{it} - \overline{y}_{i.} = (\mathbf{x}_{it} - \overline{\mathbf{x}}_{i.})'\boldsymbol{\beta} + \gamma(y_{i,t-1} - \overline{y}_{i.}) + (\varepsilon_{it} - \overline{\varepsilon}_{i.}).$$

Anderson and Hsiao (1981, 1982) show that

$$\begin{aligned} \text{Rom } \mathit{Cov}[(y_{it} - \overline{y}_{i.}), (\varepsilon_{it} - \overline{\varepsilon}_{i.})] &\approx \frac{-\sigma_\varepsilon^2}{T(1-\gamma)^2}\left[\frac{(T-1) - T\gamma + \gamma^T}{T}\right] \\ &= \frac{-\sigma_\varepsilon^2}{T(1-\gamma)^2}\left[(1-\gamma) - \frac{1-\gamma^T}{T}\right]. \end{aligned} \qquad \cancel{(9\text{-}70)}$$

11-68

This does converge to zero as T increases, but, again, we are considering cases in which T is small or moderate, say 5 to 15, in which case, the bias in the OLS estimator could be 15 percent to 60 percent. The implication is that the "within" transformation does not produce a consistent estimator.

It is easy to see that taking first differences is likewise ineffective. The first differences of the observations are

65

$$y_{it} - y_{i,t-1} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})'\boldsymbol{\beta} + \gamma(y_{i,t-1} - y_{i,t-2}) + (\varepsilon_{it} - \varepsilon_{i,t-1}). \qquad 11\text{-}62\ \cancel{(9\text{-}71)}$$

As before, the correlation between the last regressor and the disturbance persists, so OLS or GLS based on first differences would also be inconsistent. There is another approach. Write the regression in differenced form as

$$\Delta y_{it} = \Delta\mathbf{x}_{it}'\boldsymbol{\beta} + \gamma\Delta y_{i,t-1} + \Delta\varepsilon_{it}$$

or, defining $\mathbf{x}_{it}^* = [\Delta\mathbf{x}_{it}, \Delta y_{i,t-1}]$, $\varepsilon_{it}^* = \Delta\varepsilon_{it}$ and $\theta = [\boldsymbol{\beta}', \gamma]'$

$$y_{it}^* = \mathbf{x}_{it}^{*\prime}\boldsymbol{\theta} + \varepsilon_{it}^*.$$

For the pooled sample, beginning with $t = 3$, write this as

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\theta} + \boldsymbol{\varepsilon}^*.$$

The least squares estimator based on the first differenced data is

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \left[\frac{1}{n(T-3)}\mathbf{X}^{*\prime}\mathbf{X}^*\right]^{-1}\left(\frac{1}{n(T-3)}\mathbf{X}^{*\prime}\mathbf{y}^*\right)\\ &= \boldsymbol{\theta} + \left[\frac{1}{n(T-3)}\mathbf{X}^{*\prime}\mathbf{X}^*\right]^{-1}\left(\frac{1}{n(T-3)}\mathbf{X}^{*\prime}\boldsymbol{\varepsilon}^*\right).\end{aligned}$$

Assuming that the inverse matrix in brackets converges to a positive definite matrix—that remains to be shown—the inconsistency in this estimator arises because the vector in parentheses does not converge to zero. The last element is $\text{plim}_{n\to\infty}[1/(n(T-3))]$ $\Sigma_{i=1}^n\Sigma_{t=3}^T(y_{i,t-1} - y_{i,t-2})(\varepsilon_{it} - \varepsilon_{i,t-1})$ which is not zero.

Suppose there were a variable z^* such that $\text{plim}[1/(n(T-3))]\mathbf{z}^{*\prime}\boldsymbol{\varepsilon}^* = 0$ and $\text{plim}[1/(n(T-3))]\mathbf{z}^{*\prime}\mathbf{X}^* \neq \mathbf{0}$. Let $\mathbf{Z} = [\Delta\mathbf{X}, \mathbf{z}^*]$; z_{it}^* replaces $\Delta y_{i,t-1}$ in $\mathbf{x}_{it}^*$. By this construction, it appears we have a consistent estimator. Consider

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\text{IV}} &= (\mathbf{Z}'\mathbf{X}^*)^{-1}\mathbf{Z}'\mathbf{y}^*.\\ &= (\mathbf{Z}'\mathbf{X}^*)^{-1}\mathbf{Z}'(\mathbf{X}^*\boldsymbol{\theta} + \boldsymbol{\varepsilon}^*)\\ &= \boldsymbol{\theta} + (\mathbf{Z}'\mathbf{X}^*)^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}^*.\end{aligned}$$

Then, after multiplying throughout by $1/(n(T-3))$ as before, we find

$$\text{Plim}\,\hat{\boldsymbol{\theta}}_{\text{IV}} = \boldsymbol{\theta} + \text{plim}\{[1/(n(T-3))](\mathbf{Z}'\mathbf{X}^*)\}^{-1} \times \mathbf{0},$$

KT which seems to solve the problem of consistent estimation. KT

The variable z^* is an **instrumental variable,** and the estimator is an **instrumental**
KT **variable estimator** (hence the subscript on the preceding estimator). Finding suitable, valid instruments, that is, variables that satisfy the necessary assumptions, for models in which the right-hand variables are correlated with omitted factors is often challenging. In this setting, there is a natural candidate—in fact, there are several. From ~~(9-71)~~ 11-62 65, we

have at period $t = 3$

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

We could use y_{i1} as the needed variable, because it is not correlated $\varepsilon_{i3} - \varepsilon_{i2}$. Continuing in this fashion, we see that for $t = 3, 4, \ldots, T$, $y_{i,t-2}$ appears to satisfy our requirements. Alternatively, beginning from period $t = 4$, we can see that $z_{it} = (y_{i,t-2} - y_{i,t-3})$ once again satisfies our requirements. This is Anderson and Hsiao's (1981) result for instrumental variable estimation of the dynamic panel data model. It now becomes a question of which approach, levels ($y_{i,t-2}$, $t = 3, \ldots, T$), or differences ($y_{i,t-2} - y_{i,t-3}$, $t = 4, \ldots, T$) is a preferable approach. Arellano (1989) and Kiviet (1995) obtains results that suggest that the estimator based on levels is more efficient.

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

9.10 SUMMARY AND CONCLUSIONS

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

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

Key Terms and Concepts

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