10
SYSTEMS OF EQUATIONS

10.1 INTRODUCTION

There are many settings in which the single equation models of the previous chapters apply to a group of related variables. In these contexts, it makes sense to consider the several models jointly. Some examples follow.

1. Munnell's (1990) model for output by the 48 continental U.S. states is

$$\ln GSP_{it} = \beta_{1i} + \beta_{2i} \ln pc_{it} + \beta_{3i} \ln hwy_{it} + \beta_{4i} \ln water_{it} + \beta_{5i} \ln util_{it} + \beta_{6i} \ln emp_{it} + \beta_{7i}\, unemp_{it} + \varepsilon_{it}.$$

   Taken one state at a time, this provides a set of 48 linear regression models. The application develops a model in which the observations are correlated across time within a state. An important question pursued here and in the applications in the next example is whether it is valid to assume that the coefficient vector is the same for all states (individuals) in the sample.

2. The capital asset pricing model of finance specifies that for a given security,

$$r_{it} - r_{ft} = \alpha_i + \beta_i (r_{mt} - r_{ft}) + \varepsilon_{it},$$

   where r_{it} is the return over period t on security i, r_{ft} is the return on a risk-free security, r_{mt} is the market return, and β_i is the security's beta coefficient. The disturbances are obviously correlated across securities. The knowledge that the return on security i exceeds the risk-free rate by a given amount gives some information about the excess return of security j, at least for some j's. It will be useful to estimate the equations jointly rather than ignore this connection.

3. Pesaran and Smith (1995) proposed a dynamic model for wage determination in 38 UK industries. The central equation is of the form

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}_i + \gamma_i y_{i,t-1} + \varepsilon_{it}.$$

   Nair-Reichert and Weinhold's (2001) cross-country analysis of growth of developing countries takes the same form. In both cases, each group (industry, country) could be analyzed separately. However, the connections across groups and the interesting question of "poolability"—that is, whether it is valid to assume identical coefficients—is a central part of the analysis. The lagged dependent variable in the model produces a substantial complication.

4. In a model of production, the optimization conditions of economic theory imply that if a firm faces a set of factor prices $\mathbf{p}$, then its set of cost-minimizing factor demands for producing output Q will be a set of equations of the form $x_m = f_m(Q, \mathbf{p})$. The empirical model is

$$
\begin{aligned}
x_1 &= f_1(Q, \mathbf{p} \mid \boldsymbol{\theta}) + \varepsilon_1,\\
x_2 &= f_2(Q, \mathbf{p} \mid \boldsymbol{\theta}) + \varepsilon_2,\\
&\cdots\\
x_M &= f_M(Q, \mathbf{p} \mid \boldsymbol{\theta}) + \varepsilon_M,
\end{aligned}
$$

where $\boldsymbol{\theta}$ is a vector of parameters that are part of the technology and εm represents errors in optimization. Once again, the disturbances should be correlated. In addition, the same parameters of the production technology will enter all the demand equations, so the set of equations have cross-equation restrictions. Estimating the equations separately will waste the information that the same set of parameters appears in all the equations. has

AU: Are Subs OK italic?

5. The essential form of a model for equilibrium in a market is

$$
\begin{aligned}
Q_{Demand} &= \alpha_1 + \alpha_2\, Price + \alpha_3\, Income + \mathbf{d}'\boldsymbol{\alpha} + \varepsilon_{Demand},\\
Q_{Supply} &= \beta_1 + \beta_2\, Price + \mathbf{s}'\boldsymbol{\beta} + \varepsilon_{Supply},\\
Q_{Equilibrium} &= Q_{Demand} = Q_{Supply},
\end{aligned}
$$

where $\mathbf{d}$ and $\mathbf{s}$ are other variables that influence the equilibrium through their impact on the demand and supply curves, respectively. This model differs from those suggested thus far because the implication of the third equation is that *Price* is not exogenous in the equation system. The equations of this model fit more appropriately in the instrumental variables framework developed in Chapter 8 than in the regression models developed in Chapters 1 to 7. The multiple equations framework developed in this chapter provides additional results for estimating "simultaneous equations models" such as this.

The multiple equations regression model developed in this chapter provides a modeling framework that can be used in many different settings. The models of production and cost developed in Section 10.5 provide the platform for the literature on empirical analysis of firm behavior. At the macroeconomic level, the "vector autoregression models" used in Chapters 21-23 are specific forms of the seemingly unrelated regressions model of Section 10.2. The simultaneous equations model presented in Section 10.6 lies behind the specification of the large variety of specifications considered in Chapter 8.

This chapter will develop the essential theory for sets of related regression equations. Section 10.2 examines the general model in which each equation has its own fixed set of parameters, and it examines efficient estimation techniques. Section 10.2.6 examines the ~~special case of the model suggested in illustrations 3 and 4 preceding,~~ the "pooled" model with identical coefficients in all equations. Production and consumer demand models are a special case of the general model in which the equations of the model obey an adding up constraint that has important implications for specification and estimation. Section 10.3 suggests extensions of the seemingly unrelated regression model to the generalized regression models with heteroscedasticity and autocorrelation that are developed in Chapter 9. Section 10.4 broadens the seemingly unrelated regressions model to nonlinear systems of equations. In Section 10.5, we examine a classic application of the seemingly unrelated regressions model that illustrates the interesting features of the current genre of demand studies in the applied literature. The seemingly unrelated regressions model is then extended to the translog specification, which forms the platform for most recent microeconomic studies of production and cost. Finally, Section 10.6 merges the results of Chapter 8 on models with endogenous variables with the development in this chapter of multiple equation systems. In Section 10.6, we will develop **simultaneous equations models**. These are systems of equations that build on the seemingly unrelated regressions model to produce equation

AU: Please clarify "illustrations 3 and 4."

KT

AU: This KT is not in chap. list

systems that include interrelationships among the dependent variables. The supply and demand model suggested in the Introduction, of equilibrium in which price and quantity in a market are jointly determined, is an application.

10.2 THE SEEMINGLY UNRELATED REGRESSIONS MODEL

All the examples suggestd in the Introduction have a common multiple equation structure, which we may write as

$$\begin{aligned}
\mathbf{y}_1 &= \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1,\\
\mathbf{y}_2 &= \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2,\\
&\cdots\\
\mathbf{y}_M &= \mathbf{X}_M\boldsymbol{\beta}_M + \boldsymbol{\varepsilon}_M.
\end{aligned} \qquad \textbf{(10-1)}$$

There are M equations and T observations in the sample of data used to estimate them.[1] The second and third examples embody different types of constraints across equations and different structures of the disturbances. A basic set of principles will apply to them all, however.[2] The seemingly unrelated regressions (SUR) model in (10-1) is

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, M. \qquad (10\text{-}2)$$

[1] The use of T is not meant to imply any necessary connection to time series. For instance, in the fourth example above, the data might be cross-sectional.

[2] See the surveys by Srivastava and Dwivedi (1979), Srivastava and Giles (1987), and Fiebig (2001).

Define the $MT \times 1$ vector of disturbances, ~~where~~

$$\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2', \ldots, \boldsymbol{\varepsilon}_M']'.$$

We assume strict exogeneity of $\mathbf{X}_i$,

$$E[\boldsymbol{\varepsilon} \mid \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M] = \mathbf{0},$$

and homoscedasticity

$$E[\boldsymbol{\varepsilon}_m \boldsymbol{\varepsilon}_m' \mid \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M] = \sigma_{mm}\mathbf{I}_T.$$

FN 3

We assume that a total of T observations are used in estimating the parameters of the M equations.[3] Each equation involves K_i regressors, for a total of $K = \sum_{i=1}^{M} K_i$. We will require $T > K_i$. The data are assumed to be well behaved, as described in Section 4.9.1, and we shall not treat the issue separately here. For the present, we also assume that disturbances are uncorrelated across observations but correlated across equations. Therefore,

$$E[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M] = \sigma_{ij}, \quad \text{if } t = s \text{ and 0 otherwise.}$$

The disturbance formulation is, therefore,

$$E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j' \mid \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M] = \sigma_{ij}\mathbf{I}_T,$$

or

$$E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M] = \boldsymbol{\Omega} = \begin{bmatrix} \sigma_{11}\mathbf{I} & \sigma_{12}\mathbf{I} & \cdots & \sigma_{1M}\mathbf{I} \\ \sigma_{21}\mathbf{I} & \sigma_{22}\mathbf{I} & \cdots & \sigma_{2M}\mathbf{I} \\ & \vdots & & \\ \sigma_{M1}\mathbf{I} & \sigma_{M2}\mathbf{I} & \cdots & \sigma_{MM}\mathbf{I} \end{bmatrix}. \tag{10-3}$$

~~Note that when the data matrices are group-specific observations on the same variables, as in Example 10.1, the specification of this model is precisely that of the covariance structures model of Section 10.2.8 save for the extension here that allows the parameter vector to vary across groups. The covariance structures model is, therefore, a testable special case.[4]~~

It will be convenient in the discussion below to have a term for the particular kind of model in which the data matrices are group specific data sets on the same set of variables. The Grunfeld model noted in the introduction is such a case. This special case of the seemingly unrelated regressions model is a **multivariate regression model.** In contrast, the cost function model examined in Section 10.4.1 is not of this type—it consists of a cost function that involves output and prices and a set of cost share equations that have only a set of constant terms. We emphasize, this is merely a convenient term for a specific form of the SUR model, not a modification of the model itself. KT

[3] There are a few results for unequal numbers of observations, such as Schmidt (1977), Baltagi, Garvin, and Kerman (1989), Conniffe (1985), Hwang (1990), and Im (1994). But, the case of fixed T is the norm in practice.

~~[4] This is the test of "Aggregation Bias" that is the subject of Zellner (1962, 1963). (The bias results if parameter equality is incorrectly assumed.) We will examine this issue in detail in Section 10.2.4.~~

10.2.1 GENERALIZED LEAST SQUARES

Each equation is, by itself, a classical regression. Therefore, the parameters could be estimated consistently, if not efficiently, one equation at a time by ordinary least squares. The **generalized regression model** applies to the stacked model,

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{10-4}$$

Therefore, the efficient estimator is generalized least squares.[4] The model has a particularly convenient form. For the tth observation, the $M \times M$ covariance matrix of the disturbances is

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2M} \\ & \vdots & & \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM} \end{bmatrix}, \tag{10-5}$$

so, in (10-3),

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \mathbf{I} \tag{10-6}$$

and

$$\boldsymbol{\Omega}^{-1} = \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}.$$[5]

Denoting the ijth element of $\boldsymbol{\Sigma}^{-1}$ by σ^{ij}, we find that the GLS estimator is

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} = [\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{X}]^{-1}\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{y}. \tag{10-7}$$

Expanding the **Kronecker products** produces

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \sigma^{11}\mathbf{X}_1'\mathbf{X}_1 & \sigma^{12}\mathbf{X}_1'\mathbf{X}_2 & \cdots & \sigma^{1M}\mathbf{X}_1'\mathbf{X}_M \\ \sigma^{21}\mathbf{X}_2'\mathbf{X}_1 & \sigma^{22}\mathbf{X}_2'\mathbf{X}_2 & \cdots & \sigma^{2M}\mathbf{X}_2'\mathbf{X}_M \\ & & \vdots & \\ \sigma^{M1}\mathbf{X}_M'\mathbf{X}_1 & \sigma^{M2}\mathbf{X}_M'\mathbf{X}_2 & \cdots & \sigma^{MM}\mathbf{X}_M'\mathbf{X}_M \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{M} \sigma^{1j}\mathbf{X}_1'\mathbf{y}_j \\ \sum_{j=1}^{M} \sigma^{2j}\mathbf{X}_2'\mathbf{y}_j \\ \vdots \\ \sum_{j=1}^{M} \sigma^{Mj}\mathbf{X}_M'\mathbf{y}_j \end{bmatrix}.$$

The asymptotic covariance matrix for the GLS estimator is the bracketed inverse matrix in (10-7). All the results of Chapter 8 for the generalized regression model extend to this model (which has both **heteroscedasticity** and **autocorrelation**).

This estimator is obviously different from ordinary least squares. At this point, however, the equations are linked only by their disturbances—hence the name *seemingly unrelated* regressions model—so it is interesting to ask just how much efficiency is gained by using generalized least squares instead of ordinary least squares. Zellner (1962) and Dwivedi and Srivastava (1978) have analyzed some special cases in detail.

[4] See Zellner (1962) and Telser (1964).

[5] See Appendix Section A.5.5.

10-6

1. If the equations are *actually* unrelated—that is, if $\sigma_{ij} = 0$ for $i \neq j$—then there is obviously no payoff to GLS estimation of the full set of equations. Indeed, full GLS is equation by equation OLS.[6]
2. If the equations have **identical explanatory variables**—that is, if $\mathbf{X}_i = \mathbf{X}_j$—then OLS and GLS are identical. We will turn to this case in Section 10.2.2.[7]
3. If the regressors in one block of equations are a subset of those in another, then GLS brings no efficiency gain over OLS in estimation of the smaller set of equations; thus, GLS and OLS are once again identical. We will look at an application of this result in Section 20.6.5.[8]

21

In the more general case, with unrestricted correlation of the disturbances and different regressors in the equations, the results are complicated and dependent on the data. Two propositions that apply generally are as follows:

1. The greater is the correlation of the disturbances, the greater is the efficiency gain accruing to GLS.
2. The less correlation there is between the $\mathbf{X}$ matrices, the greater is the gain in efficiency in using GLS.[9]

10.2.2 SEEMINGLY UNRELATED REGRESSIONS WITH IDENTICAL REGRESSORS

21

The case of **identical regressors** is quite common, notably in the capital asset pricing model in empirical finance—see the Introduction and Chapter 20. In this special case, generalized least squares is equivalent to equation by equation ordinary least squares. Impose the assumption that $\mathbf{X}_i = \mathbf{X}_j = \mathbf{X}$, so that $\mathbf{X}_i'\mathbf{X}_j = \mathbf{X}'\mathbf{X}$ for all i and j in (10-7). The inverse matrix on the right-hand side now becomes $[\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}'\mathbf{X}]^{-1}$, which, using (A-76), equals $[\boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}]$. Also on the right-hand side, each term $\mathbf{X}_i'\mathbf{y}_j$ equals $\mathbf{X}'\mathbf{y}_j$, which, in turn equals $\mathbf{X}'\mathbf{X}\mathbf{b}_j$. With these results, after moving the common $\mathbf{X}'\mathbf{X}$ out of the summations on the right-hand side, we obtain

chapter/

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \sigma_{11}(\mathbf{X}'\mathbf{X})^{-1} & \sigma_{12}(\mathbf{X}'\mathbf{X})^{-1} & \cdots & \sigma_{1M}(\mathbf{X}'\mathbf{X})^{-1} \\ \sigma_{21}(\mathbf{X}'\mathbf{X})^{-1} & \sigma_{22}(\mathbf{X}'\mathbf{X})^{-1} & \cdots & \sigma_{2M}(\mathbf{X}'\mathbf{X})^{-1} \\ & & \vdots & \\ \sigma_{M1}(\mathbf{X}'\mathbf{X})^{-1} & \sigma_{M2}(\mathbf{X}'\mathbf{X})^{-1} & \cdots & \sigma_{MM}(\mathbf{X}'\mathbf{X})^{-1} \end{bmatrix} \begin{bmatrix} (\mathbf{X}'\mathbf{X})\sum_{l=1}^{M}\sigma^{1l}\mathbf{b}_l \\ (\mathbf{X}'\mathbf{X})\sum_{l=1}^{M}\sigma^{2l}\mathbf{b}_l \\ \vdots \\ (\mathbf{X}'\mathbf{X})\sum_{l=1}^{M}\sigma^{Ml}\mathbf{b}_l \end{bmatrix}. \quad \text{(10-8)}$$

[6]See also Baltagi (1989) and Bartels and Fiebig (1992) for other cases in which OLS = GLS.

[7]An intriguing result, albeit probably of negligible practical significance, is that the result also applies if the $\mathbf{X}$'s are all nonsingular, and not necessarily identical, linear combinations of the same set of variables. The formal result which is a corollary of Kruskal's theorem [see Davidson and MacKinnon (1993, p. 294)] is that OLS and GLS will be the same if the K columns of $\mathbf{X}$ are a linear combination of exactly K characteristic vectors of $\boldsymbol{\Omega}$. By showing the equality of OLS and GLS here, we have verified the conditions of the corollary. The general result is pursued in the exercises. The intriguing result cited is now an obvious case.

[8]The result was analyzed by Goldberger (1970) and later by Revankar (1974) and Conniffe (1982a, b).

[9]See also Binkley (1982) and Binkley and Nelson (1988).

10-7

Now, we isolate one of the subvectors, say the first, from $\hat{\boldsymbol{\beta}}$. After multiplication, the moment matrices cancel, and we are left with

$$\hat{\boldsymbol{\beta}}_1 = \sum_{j=1}^{M} \sigma_{1j} \sum_{l=1}^{M} \sigma^{jl} \mathbf{b}_l = \mathbf{b}_1 \left(\sum_{j=1}^{M} \sigma_{1j} \sigma^{j1} \right) + \mathbf{b}_2 \left(\sum_{j=1}^{M} \sigma_{1j} \sigma^{j2} \right) + \cdots + \mathbf{b}_M \left(\sum_{j=1}^{M} \sigma_{1j} \sigma^{jM} \right).$$

The terms in parentheses are the elements of the first row of $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1} = \mathbf{I}$, so the end result is $\hat{\boldsymbol{\beta}}_1 = \mathbf{b}_1$. For the remaining subvectors, which are obtained the same way, $\hat{\boldsymbol{\beta}}_i = \mathbf{b}_i$, which is the result we sought.[10]

FN 10

To reiterate, the important result we have here is that in the SUR model, when all equations have the same regressors, the efficient estimator is single-equation ordinary least squares; OLS is the same as GLS. Also, the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ for this case is given by the large inverse matrix in brackets in (10-8), which would be estimated by

$$\text{Est. Asy. Cov}[\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\beta}}_j] = \hat{\sigma}_{ij}(\mathbf{X}'\mathbf{X})^{-1}, \quad i, j = 1, \ldots, M, \quad \text{where } \hat{\boldsymbol{\Sigma}}_{ij} = \hat{\sigma}_{ij} = \frac{1}{T}\mathbf{e}_i'\mathbf{e}_j.$$

Except in some special cases, this general result is lost if there are any restrictions on $\boldsymbol{\beta}$, either within or across equations. We will examine one of those cases, the block of zeros restriction, in Section ~~20~~.6.5.

21

10.2.3 FEASIBLE GENERALIZED LEAST SQUARES

FN 11

The preceding discussion assumes that $\boldsymbol{\Sigma}$ is known, which, as usual, is unlikely to be the case. FGLS estimators have been devised, however.[11] The least squares residuals may be used (of course) to estimate consistently the elements of $\boldsymbol{\Sigma}$ with

$$\hat{\sigma}_{ij} = s_{ij} = \frac{\mathbf{e}_i'\mathbf{e}_j}{T}. \tag{10-9}$$

FN 12

The consistency of s_{ij} follows from that of $\mathbf{b}_i$ and $\mathbf{b}_j$.[12] A degrees of freedom correction in the divisor is occasionally suggested. Two possibilities that are unbiased when $i = j$ are

FN 13

$$s_{ij}^* = \frac{\mathbf{e}_i'\mathbf{e}_j}{[(T - K_i)(T - K_j)]^{1/2}} \quad \text{and} \quad s_{ij}^{**} = \frac{\mathbf{e}_i'\mathbf{e}_j}{T - \max(K_i, K_j)}.^{13} \tag{10-10}$$

Whether unbiasedness of the estimator of $\boldsymbol{\Sigma}$ used for FGLS is a virtue here is uncertain. KT The asymptotic properties of the **feasible GLS** estimator, $\hat{\hat{\boldsymbol{\beta}}}$ do not rely on an unbiased estimator of $\boldsymbol{\Sigma}$; only consistency is required. All our results from Chapters 8 and 9 for FGLS estimators extend to this model, with no modification. We shall use (10-9)

AQ: Term "Zellner's efficient estimator" is not in chap. list

[10] See Hashimoto and Ohtani (1990) for discussion of hypothesis testing in this case.

[11] See Zellner (1962) and Zellner and Huang (1962). The FGLS estimator for this model is also labeled KT **Zellner's efficient estimator**, or ZEF, in reference to Zellner (1962) where it was introduced.

[12] Perhaps surprisingly, if it is assumed that the density of $\boldsymbol{\varepsilon}$ is symmetric, as it would be with normality, then $\mathbf{b}_i$ is also unbiased. See Kakwani (1967).

[13] See, as well, Judge et al. (1985), Theil (1971), and Srivastava and Giles (1987).

10-8

in what follows. With

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1M} \\ s_{21} & s_{22} & \cdots & s_{2M} \\ & \vdots & & \\ s_{M1} & s_{M2} & \cdots & s_{MM} \end{bmatrix} \tag{10-11}$$

in hand, FGLS can proceed as usual.

Insert Example 10.X here

Example 10.2 Estimated SUR Model for Regional Output

We initially estimated the nine equations of the regional productivity model separately by OLS. The OLS estimates are shown in Table 10.1. The correlation matrix for the OLS residuals is as follows:

replace

R =	GF	MW	MA	MT	NE	SO	SW	CN	WC
GF	1.0000								
MW	−0.1782	1.0000							
MA	0.0816	0.8362	1.0000						
MT	0.6418	0.4746	0.6837	1.0000					
NE	−0.0250	0.0797	−0.1736	−0.1873	1.0000				
SO	0.0654	−0.4630	−0.0436	−0.1190	−0.5338	1.0000			
SW	−0.1032	−0.4845	−0.1709	−0.3502	−0.3041	0.9339	1.0000		
CN	0.6194	−0.8034	−0.4376	0.1160	−0.1924	0.4745	0.3507	1.0000	
WC	0.6759	−0.1774	0.2312	0.6723	−0.1261	0.3124	0.1525	0.6845	1.0000

The values in **R** are large enough to suggest that there is substantial correlation of the disturbances across regions. As noted in Section 10.2.7, we do not have a standard test of the specification of the SUR model against the alternative hypothesis of uncorrelated disturbances for the general SUR model without an assumption of normality. The Breusch and Pagan (1980) Lagrange multiplier test based on the correlation matrix does have some intuitive appeal. For the correlation matrix shown earlier, the chi-squared statistic equals 102.305 ~~215.88~~ with $8(9)/2 = 36$ degrees of freedom. The critical value from the chi-squared table is 50.998, so the null hypothesis that the seemingly unrelated regressions are actually unrelated is rejected.

Table 10.1 presents estimates of the parameters of the SUR model for regional output. The table lists the OLS as well as the FGLS parameter estimates. (The pooled results that are also presented are discussed in Section 10.2.8.) The correlations listed earlier suggest that there is likely to be considerable benefit to using FGLS in terms of efficiency of the estimator. The individual equation OLS estimators are consistent, but they neglect the cross equation correlation. The substantially lower estimated standard errors for the FGLS results with each equation appear to confirm that expectation.

explain later fgls

10.2.4 TESTING HYPOTHESES

For testing a hypothesis about $\boldsymbol{\beta}$, a statistic analogous to the F ratio in multiple regression analysis is

delete 1 over Ω

$$F[J, MT-K] = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})/J}{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\varepsilon}}/(MT-K)}. \tag{10-12}$$

The computation requires the unknown $\boldsymbol{\Omega}$. If we insert the FGLS estimate $\hat{\boldsymbol{\Omega}}$ based on (10-9) and use the result that the denominator converges to one, then, in large samples,

10-9

the statistic will behave the same as

$$\hat{F} = \frac{1}{J}(\mathbf{R}\hat{\hat{\beta}} - \mathbf{q})'[\mathbf{R}\,\widehat{\mathrm{Var}}[\hat{\hat{\beta}}]\mathbf{R}']^{-1}(\mathbf{R}\hat{\hat{\beta}} - \mathbf{q}). \tag{10-13}$$

This can be referred to the standard F table. Because it uses the estimated $\boldsymbol{\Sigma}$, even with normally distributed disturbances, the F distribution is only valid approximately. In general, the statistic $F[J, n]$ converges to $1/J$ times a chi-squared $[J]$ as $n \to \infty$. Therefore, an alternative test statistic that has a limiting chi-squared distribution with J degrees of freedom when the null hypothesis is true is

$$J\hat{F} = (\mathbf{R}\hat{\hat{\beta}} - \mathbf{q})'[\mathbf{R}\widehat{\mathrm{Var}}[\hat{\hat{\beta}}]\mathbf{R}']^{-1}(\mathbf{R}\hat{\hat{\beta}} - \mathbf{q}). \tag{10-14}$$

AQ: Term "Wald statistic" is not in chap. list

KT

This can be recognized as a **Wald statistic** that measures the distance between $\mathbf{R}\hat{\hat{\beta}}$ and $\mathbf{q}$. Both statistics are valid asymptotically, but (10-13) may perform better in a small or moderately sized sample.[14] Once again, the divisor used in computing $\hat{\sigma}_{ij}$ may make a difference, but there is no general rule.

FN 14

KT

A hypothesis of particular interest is the **homogeneity restriction** of equal coefficient vectors in the multivariate regression model. That case is fairly common in this setting. The homogeneity restriction is that $\boldsymbol{\beta}_i = \boldsymbol{\beta}_M, i = 1, \dots, M-1$. Consistent with (10-13)–(10-14), we would form the hypothesis as

$$\mathbf{R}\boldsymbol{\beta} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & -\mathbf{I} \\ \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} & -\mathbf{I} \\ & & \cdots & & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & -\mathbf{I} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \cdots \\ \boldsymbol{\beta}_M \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_1 - \boldsymbol{\beta}_M \\ \boldsymbol{\beta}_2 - \boldsymbol{\beta}_M \\ \cdots \\ \boldsymbol{\beta}_{M-1} - \boldsymbol{\beta}_M \end{pmatrix} = \mathbf{0}. \tag{10-15}$$

This specifies a total of $(M-1)K$ restrictions on the $KM \times 1$ parameter vector. Denote the estimated asymptotic covariance for $(\hat{\hat{\boldsymbol{\beta}}}_i, \hat{\hat{\boldsymbol{\beta}}}_j)$ as $\hat{\mathbf{V}}_{ij}$. The bracketed matrix in (10-13) would have typical block

$$[\mathbf{R}\,\widehat{\mathrm{Var}}[\hat{\hat{\beta}}]\mathbf{R}']_{ij} = \hat{\mathbf{V}}_{ij} - \hat{\mathbf{V}}_{iM} - \hat{\mathbf{V}}_{Mj} + \hat{\mathbf{V}}_{MM}$$

14

This may be a considerable amount of computation. The test will be simpler if the model has been fit by maximum likelihood, as we examine in Section 16.9.3. Pesaran and Yamagata (2008) provide an alternative test that can be used when M is large and T is relatively small.

Example 10.3 Hypothesis Tests in the SUR Model

We used (10-14) to construct test statistics for two hypotheses. The "pooling" restriction for the multivariate regression (same variables—not necessarily the same data, as in our example) is formulated as

$$H_0: \quad \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_M,$$
$$H_1: \quad \text{Not } H_0.$$

For this hypothesis, the $\mathbf{R}$ matrix is shown in (10-15). The test statistic is in (10-14). For our model with nine equations and seven parameters in each, the null hypothesis imposes $8(7) = 56$ restrictions. The computed test statistic is 10,554.77, which is far larger than the critical value from the table, 74.468. So, the hypothesis of homogeneity is rejected.

[14] See Judge et al. (1985, p. 476). The Wald statistic often performs poorly in the small sample sizes typical in this area. Fiebig (2001, pp. 108–110) surveys a recent literature on methods of improving the power of testing procedures in SUR models.

10-10

"corrections" for autocorrelation. At one extreme is Mizon (1995) who argues forcefully that autocorrelation arises as a consequence of a remediable failure to include dynamic effects in the model. However, in a system of equations, the analysis that leads to this conclusion is going to be far more complex than in a single equation model.[15] Suffice to say, the issue remains to be settled conclusively.

10.2.5 A SPECIFICATION TEST FOR THE SUR MODEL

It is of interest to assess statistically whether the off diagonal elements of $\mathbf{\Sigma}$ are zero. If so, then the efficient estimator for the full parameter vector, absent heteroscedasticity or autocorrelation, is equation by equation ordinary least squares. There is no standard test for the general case of the SUR model unless the additional assumption of normality of the disturbances is imposed in (10-2) and (10-3). With normally distributed disturbances, the standard trio of tests, Wald, **likelihood ratio,** and **Lagrange multiplier,** can be used. For reasons we will turn to shortly, the Wald test is likely to be too cumbersome to apply. With normally distributed disturbances, the likelihood ratio statistic for testing the null hypothesis that the matrix $\mathbf{\Sigma}$ in (10-5) is a diagonal matrix against the alternative that $\mathbf{\Sigma}$ is simply an unrestricted positive definite matrix would be

KT KT

$$\lambda_{LR} = T[\ln|\mathbf{S}_0| - \ln|\mathbf{S}_1|], \tag{10-16}$$

where $\mathbf{S}_h$ is the residual covariance matrix defined in (10-9) (without a degrees of freedom correction). The residuals are computed using maximum likelihood estimates of the parameters, not FGLS.[16] Under the null hypothesis, the model would be efficiently estimated by individual equation OLS, so

FN 15 15

$$\ln|\mathbf{S}_0| = \sum_{i=1}^{M} \ln\left(\mathbf{e}_i'\mathbf{e}_i/T\right),$$

where $\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i\mathbf{b}_i$. The limiting distribution of the likelihood ratio statistic under the null hypothesis would be chi-squared with $M(M-1)/2$ degrees of freedom.

The likelihood ratio statistic requires the unrestricted MLE to compute the residual covariance matrix under the alternative, so it is can be cumbersome to compute. A simpler alternative is the Lagrange multiplier statistic developed by Breusch and

[15] Dynamic SUR models in the spirit of Mizon's admonition were proposed by Anderson and Blundell (1982). A few recent applications are Kiviet, Phillips, and Schipp (1995) and DesChamps (1998). However, relatively little work has been done with dynamic SUR models. The VAR models in Section 20.6 are an important group of applications, but they come from a different analytical framework. Likewise, the panel data applications noted in the Introduction and in Section 9.8.5 would fit into the modeling framework we are developing here. However, in these applications, the regressions are "actually" unrelated—the authors did not model the cross-unit correlation that is the central focus of this chapter. Related results may be found in Guilkey and Schmidt (1973), Guilkey (1974), Berndt and Savin (1977), Moschino and Moro (1994), McLaren (1996), and Holt (1998).

15 [16] In the SUR model of this chapter, the MLE for normally distributed disturbances can be computed by iterating the FGLS procedure, back and forth between (10-7) and (10-9) until the estimates are no longer changing. We note, this procedure produces the MLE when it converges, but it is not guaranteed to converge, nor is it assured that there is a unique MLE. For our regional data set, the iterated FGLS procedure does not converge after 1,000 iterations. The Oberhofer–Kmenta (1974) result implies that if the iteration converges, it reaches the MLE. It does not guarantee that the iteration will converge, however. The problem with this application may be the very small sample size, 17 observations. One would not normally use the technique of maximum likelihood with a sample this small.

Pagan (1980) which is

$$\lambda_{LM} = T\sum_{i=2}^{M}\sum_{j=1}^{i-1} r_{ij}^2$$
$$= (T/2)[trace(\mathbf{R}'\mathbf{R}) - M], \tag{10-17}$$

where $\mathbf{R}$ is the sample correlation matrix of the M sets of T OLS residuals. This has the same large sample distribution under the null hypothesis as the likelihood ratio statistic, but is obviously easier to compute, as it only requires the OLS residuals.

102.305

8399.91

Example 10.4 Testing for Cross-Equation Correlation

We used (10-17) to compute the LM statistic for the 9 equation model reported in Table 10.1. The chi-squared statistic is 215.879 with $9(8)/2 = 36$ degrees of freedom. The critical value is 50.998, so we conclude that the disturbances in the regional model are not actually unrelated. The null hypothesis that $\sigma_{ij} = 0$ for all $i \neq j$ is rejected. To investigate a bit further, we repeated the test with the completely disaggregated (statewide) data. The corresponding chi-squared statistic is 2455.71 with $48(47)/2 = 1{,}128$ degrees of freedom. The critical value is 1,207.25, so the null hypothesis is rejected at the state level as well.

The third test statistic in the trio is the Wald statistic. In principle, the Wald statistic for the SUR model would be computed using

$$W = \hat{\sigma}'[Asy.\ Var(\hat{\sigma})]^{-1}\hat{\sigma},$$

where $\hat{\sigma}$ is the $M(M-1)/2$ length vector containing the estimates of the off-diagonal (lower triangle) elements of $\boldsymbol{\Sigma}$, and the asymptotic covariance matrix of the estimator appears in the brackets. Under normality, the asymptotic covariance matrix contains the corresponding elements of $2\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}/T$. It would be possible to estimate the covariance term more generally using a moment-based estimator. Because

$$\hat{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T} e_{it}e_{jt}$$

is a mean of T observations, one might use the conventional estimator of its variance and its covariance with $\hat{\sigma}_{lm}$, which would be

$$f_{ij,lm} = \frac{1}{T}\frac{1}{T-1}\sum_{t=1}^{T}(e_{it}e_{jt} - \hat{\sigma}_{ij})(e_{lt}e_{mt} - \hat{\sigma}_{lm}). \tag{10-18}$$

The modified Wald statistic would then be

$$W' = \hat{\sigma}'[\mathbf{F}]^{-1}\hat{\sigma}$$

where the elements of $\mathbf{F}$ are the corresponding values in (10-18). This computation is obviously more complicated than the other two. However, it does have the virtue that it does not require an assumption of normality of the disturbances in the model. What would be required is (a) consistency of the estimators of β_i so that the we can assert (b) consistency of the estimators of σ_{ij} and, finally, (c) asymptotic normality of the estimators in (b) so that we can apply Theorem ~~4.5~~ 4.4. All three requirements should be met in the SUR model with well behaved regressors.

Alternative approaches that have been suggested [see, e.g., Johnson and Wichern (2005, p. 424)] are based on the following general strategy: Under the alternative hypothesis of an unrestricted $\boldsymbol{\Sigma}$, the sample estimate of $\boldsymbol{\Sigma}$ will be $\hat{\boldsymbol{\Sigma}} = [\hat{\sigma}_{ij}]$ as defined

10-12

in (10-9). Under any restrictive null hypothesis, the estimator of $\boldsymbol{\Sigma}$ will be $\hat{\boldsymbol{\Sigma}}_0$, a matrix that by construction will be larger than $\hat{\boldsymbol{\Sigma}}_1$ in the matrix sense defined in Appendix A. Statistics based on the "excess variation," such as $T(\hat{\boldsymbol{\Sigma}}_0 - \hat{\boldsymbol{\Sigma}}_1)$ are suggested for the testing procedure. One of these is the likelihood ratio test in (10-16).

10.2.~~8~~6 THE POOLED MODEL

If the variables in $\mathbf{X}_i$ are all the same and the coefficient vectors in (10-2) are assumed all to be equal, the **pooled model,** KT

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

results. This differs from the panel data treatment in Chapter ~~9~~ 11, however, in that the correlation across observations is assumed to occur at time t, not within group i. (Of course, by a minor rearrangement of the data, the same model results. However, the interpretation differs, so we will maintain the distinction.) Collecting the T observations for group i, we obtain

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

or, for all n groups,

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix}\boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{10-19}$$

where

$$\begin{aligned} E[\boldsymbol{\varepsilon}_i \mid \mathbf{X}] &= \mathbf{0}, \\ E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j' \mid \mathbf{X}] &= \sigma_{ij}\boldsymbol{\Omega}_{ij}. \end{aligned} \tag{10-20}$$

If $\boldsymbol{\Omega}_{ij} = \mathbf{I}$, then this is equivalent to the SUR model of (10-2) with identical coefficient vectors. The generalized least squares estimator under this **covariance structures model** KT assumption is

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= [\mathbf{X}'(\boldsymbol{\Sigma}\otimes\mathbf{I})^{-1}\mathbf{X}]^{-1}[\mathbf{X}'(\boldsymbol{\Sigma}\otimes\mathbf{I})^{-1}\mathbf{y}] \\ &= \left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma^{ij}\mathbf{X}_i'\mathbf{X}_j\right]^{-1}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma^{ij}\mathbf{X}_i'\mathbf{y}_j\right]. \end{aligned} \tag{10-21}$$

where σ^{ij} denotes the ijth element of $\boldsymbol{\Sigma}^{-1}$. The FGLS estimator can be computed using (10-9), where $\mathbf{e}_i$ can either be computed using group-specific OLS residuals or it can be a subvector of the pooled OLS residual vector using all nT observations.

There is an important consideration to note in feasible GLS estimation of this model. The computation requires inversion of the matrix $\hat{\boldsymbol{\Sigma}}$ where the ijth element is given by (10-9). This matrix is $n \times n$. It is computed from the least squares residuals using

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{e}_t' = \frac{1}{T}\mathbf{E}'\mathbf{E}, \tag{10-22}$$

where $\mathbf{e}_t'$ is a $1 \times n$ vector containing all n residuals for the n groups at time t, placed as the tth row of the $T \times n$ matrix of residuals, $\mathbf{E}$. The rank of this matrix cannot be

10-13

larger than T. Note what happens if $n > T$. In this case, the $n \times n$ matrix has rank T, which is less than n, so it must be singular, and the FGLS estimator cannot be computed.
10.1
Consider Example 10.2. We aggregated the 48 states into $n = 9$ regions. It would not be possible to fit a full model for the $n = 48$ states with only $T = 17$ observations. This result is a deficiency of the data set, not the model. The population matrix, $\mathbf{\Sigma}$ is positive definite. But, if there are not enough observations, then the data set is too short to obtain a positive definite estimate of the matrix.

10.3 PANEL DATA APPLICATIONS

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

10.3.1 RANDOM EFFECTS SUR MODELS

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

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

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

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

and

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

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

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

Example 10.1 A Regional Production Model for Public Capital proposed a model of

~~In Examples 9.9 and 9.12, we examined~~ Munnell~~'s~~ (1990) ~~study of~~ productivity of public capital at the state level. The central equation of the analysis that we will extend here is a Cobb–Douglas production function,

$$\ln gsp_{it} = \alpha_i + \beta_{1i} \ln pc_{it} + \beta_{2i} \ln hwy_{it} + \beta_{3i} \ln water_{it} + \beta_{4i} \ln util_{it} + \beta_{5i} \ln emp_{it} + \beta_{6i}\, unemp_{it} + \varepsilon_{it},$$

where the variables in the model, measured for the lower 48 U.S. states and years 1970–1986, are

gsp	= gross state product,
pc	= public capital,
hwy	= highway capital,
water	= water utility capital,
util	= utility capital,
p_cap	= private capital,
emp	= employment (labor) ,
unemp	= unemployment rate.

In Example 9.9, we defined nine regions consisting of groups of the 48 states:

1. GF	= *Gulf*	= AL, FL, LA, MS,
2. MW	= *Midwest*	= IL, IN, KY, MI, MN, OH, WI,
3. MA	= *Mid Atlantic*	= DE, MD, NJ, NY, PA, VA,
4. MT	= *Mountain*	= CO, ID, MT, ND, SD, WY,
5. NE	= *New England*	= CT, ME, MA, NH, RI, VT,
6. SO	= *South*	= GA, NC, SC, TN, WV,R
7. SW	= *Southwest*	= AZ, NV, NM, TX, UT,
8. CN	= *Central*	= AK, IA, KS, MO, NE, OK,
9. WC	= *West Coast*	= CA, OR, WA.

For our application in this chapter, we will use the aggregated data to analyze a nine-region (equation) model. Data on output, the capital stocks, and employment are aggregated simply by summing the values for the individual states (before taking logarithms). The unemployment rate for each region, m, at time t is determined by a weighted average of the unemployment rates for the states in the region, where the weights are

$$w_{it} = emp_{it} / \Sigma_j\, emp_{jt}.$$

Then, the unemployment rate for region m at time t is the following average of the unemployment rates of the states (j) in region (m) at time t:

$$unemp_{mt} = \Sigma_j\, w_{jt}(m)\, unemp_{jt}(m).$$

TB 10.1

We initially estimated the nine equations of the regional productivity model separately by OLS. The OLS estimates are shown in Table 10.1. The correlation matrix for the OLS residuals is as follows:

10-15

		GF	MW	MA	MT	NE	SO	SW	CN	WC
	GF	1.0000								
	MW	0.1036	1.0000							
	MA	0.3421	0.0634	1.0000						
R =	MT	0.4243	0.6970	−0.0158	1.0000					
	NE	−0.5127	−0.2896	0.1915	−0.5372	1.0000				
	SO	0.5897	0.4893	0.2329	0.3434	−0.2411	1.0000			
	SW	0.3115	0.1320	0.6514	0.1301	−0.3220	0.2594	1.0000		
	CN	0.7958	0.3370	0.3904	0.4957	−02980	0.8050	0.3465	1.0000	
	WC	0.2340	0.5654	0.2116	0.5736	−0.0576	0.2693	−0.0375	0.3818	1.0000

Note minus signs

The values in **R** are large enough to suggest that there is substantial correlation of the disturbances across regions.

Table 10.1 also presents the FGLS estimates of the parameters of the SUR model for regional output. These are computed in two steps, with the first step OLS results producing the estimate of Σ for FGLS. (The pooled results that are also presented are discussed in Section 10.2.8.) The correlations listed earlier suggest that there is likely to be considerable benefit to using FGLS in terms of efficiency of the estimator. The individual equation OLS estimators are consistent, but they neglect the cross equation correlation. The substantially lower estimated standard errors for the FGLS results with each equation appear to confirm that expectation.

We used (10-14) to construct test statistics for two hypotheses. We first tested the hypothesis of constant returns to scale throughout the system. Constant returns to scale would require that the coefficients on the inputs, β_2 through β_6 (four capital variables and the labor variable) sum to 1.0. The $9\times 9(7)$ matrix, **R**, for (10-14) would have rows equal to

(mult)

$$\mathbf{R}_1 = (0,1,1,1,1,1,0) \;\; \mathbf{0}' \;\; \mathbf{0}' \;\; \mathbf{0}' \;\; \mathbf{0}' \;\; \mathbf{0}' \;\; \mathbf{0}' \;\; \mathbf{0}' \;\; \mathbf{0}'$$
$$\mathbf{R}_2 = \mathbf{0}' \;\; (0,1,1,1,1,1,0) \;\; \mathbf{0}' \;\; \mathbf{0}' \;\; \mathbf{0}' \;\; \mathbf{0}' \;\; \mathbf{0}' \;\; \mathbf{0}' \;\; \mathbf{0}'$$

and so on. In (10-14), we would have $\mathbf{q}' = (1,1,1,1,1,1,1,1,1)$. This hypothesis imposes nine restrictions. The computed chi-squared is 102.305. The critical value is 16.919, so this hypothesis is rejected as well. The discrepancy vector for these results is

$$(\mathbf{R}\hat{\boldsymbol{\beta}}-\mathbf{q})' = (-0.64674,-0.12883,0.96435,0.03930,0.06710,1.79472,2.30283,.12907,1.10534).$$

The distance is quite large for some regions, so the hypothesis of constant returns to scale (to the extent it is meaningful at this level of aggregation) does appear to be inconsistent with the data (results).

The "pooling" restriction for the multivariate regression (same variables—not necessarily the same data, as in our example) is formulated as

$$H_0: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_M,$$
$$H_1: \text{Not } H_0.$$

For this hypothesis, the **R** matrix is shown in (10-15). The test statistic is in (10-14). For our model with nine equations and seven parameters in each, the null hypothesis imposes 8(7) = 56 restrictions. The computed test statistic is 10,554.77, which is far larger than the critical value from the table, 74.468. So, the hypothesis of homogeneity is rejected.

As noted in Section 10.2.7, we do not have a standard test of the specification of the SUR model against the alternative hypothesis of uncorrelated disturbances for the general SUR model without an assumption of normality. The Breusch and Pagan (1980) Lagrange multiplier test based on the correlation matrix does have some intuitive appeal. We used (10-17) to compute the LM statistic for the 9 equation model reported in Table 10.1. For the correlation matrix shown earlier, the chi-squared statistic equals 102.305 with 8(9)/2 = 36 degrees of freedom. The critical value from the chi-squared table is 50.998, so the null hypothesis that the seemingly unrelated regressions are actually unrelated is rejected. We conclude that the disturbances in the regional model are not actually unrelated. The null hypothesis that $\sigma_{ij} = 0$ for nine

10-16

all $i \neq j$ is rejected. To investigate a bit further, we repeated the test with the completely disaggregated (statewide) data. The corresponding chi-squared statistic is 8399.41 with 48(47)/2 = 1,128 degrees of freedom. The critical value is 1,207.25, so the null hypothesis is rejected at the state level as well.

10-17

TABLE 10.1 Estimated SUR Model for Regional Output. (standard errors in parentheses)

Region	Estimator	α	β_1	β_2	β_3	β_4	β_5	β_6	σ_m	R^2
GF	OLS	12.1458	−0.007117	−2.1352	0.1161	1.4247	0.7851	−0.00742	0.01075	0.9971
		(3.3154)	(0.01114)	(0.8677)	(0.06278)	(0.5944)	(0.1493)	(0.00316)		
	FGLS	10.4792	−0.003160	−1.5448	0.1139	0.8987	0.8886	−0.005299	0.008745	0.9967
		(1.5912)	(0.005391)	(0.3888)	(0.03651)	(0.2516)	(0.07715)	(0.00182)		
MW	OLS	3.0282	0.1635	−0.07471	−0.1689	0.6372	0.3622	−0.01736	0.009942	0.9984
		(1.7834)	(0.1660)	(0.2205)	(0.09896)	(0.2078)	(0.1650)	(0.004741)		
	FGLS	4.1206	0.06370	−0.1275	−0.1292	0.5144	0.5497	−0.01545	0.008608	0.9980
		(1.0091)	(0.08739)	(0.1284)	(0.06152)	(0.1118)	(0.08597)	(0.00252)		
MA	OLS	−11.2110	0.4120	2.1355	0.5122	−0.4740	−0.4620	−0.03022	0.01040	0.9950
		(3.5867)	(0.2281)	(0.5571)	(0.1192)	(0.2519)	(0.3529)	(0.00853)		
	FGLS	−9.1438	0.3511	1.7972	0.5168	−0.3616	−0.3391	−0.02954	0.008625	0.9946
		(2.2025)	(0.1077)	(0.3410)	(0.06405)	(0.1294)	(0.1997)	(0.00474)		
MT	OLS	3.5902	0.2948	0.1740	−0.2257	−0.2144	0.9166	−0.008143	0.01688	0.9940
		(6.9490)	(0.2054)	(0.2082)	(0.3840)	(0.9712)	(0.3772)	(0.00839)		
	FGLS	2.8150	0.1843	0.1164	−0.3811	0.01648	1.1032	−0.005507	0.01321	0.9938
		(3.4428)	(0.09220)	(0.1165)	(0.1774)	(0.4654)	(0.1718)	(0.00422)		
NE	OLS	6.3783	−0.1526	−0.1233	0.3065	−0.5326	1.3437	0.005098	0.008601	0.9986
		(2.3823)	(0.08403)	(0.2850)	(0.08917)	(0.2375)	(0.1876)	(0.00517)		
	FGLS	3.5331	−0.1097	0.1637	0.2459	−0.3155	1.0828	−0.000664	0.007249	0.9983
		(1.3388)	(0.04570)	(0.1676)	(0.04974)	(0.1194)	(0.09248)	(0.00263)		
SO	OLS	−13.7297	−0.02040	0.6621	−0.9693	−0.1074	3.3803	0.03378	0.02241	0.9851
		(18.0199)	(0.2856)	(1.8111)	(0.2843)	(0.5634)	(1.1643)	(0.02150)		
	FGLS	−13.1186	0.1007	0.9923	−0.5851	−0.3029	2.5897	0.02143	0.01908	0.9817
		(7.6009)	(0.1280)	(0.7827)	(0.1373)	(0.2412)	(0.4665)	(0.00809)		

Region	Estimator	α	β_1	β_2	β_3	β_4	β_5	β_6	σ_m	R^2
SW	***OLS***	−22.8553 (4.8739)	−0.3776 (0.1673)	3.3478 (1.8584)	−0.2637 (0.4317)	−1.7783 (1.1757)	2.6732 (1.0325)	0.02592 (0.01727)	0.01293	0.9864
	FGLS	−19.9917 (2.8649)	−0.3386 (0.08943)	3.2821 (0.8894)	−0.1105 (0.1993)	−1.7812 (0.5609)	2.2510 (0.4802)	0.01846 (0.00793)	0.01055	0.9846
CN	***OLS***	3.4425 (1.2571)	0.05040 (0.2662)	−0.5938 (0.3219)	0.06351 (0.3333)	−0.01294 (0.3787)	1.5731 (0.4125)	0.006125 (0.00892)	0.01753	0.9936
	FGLS	2.8172 (0.8434)	0.01412 (0.08833)	−0.5086 (0.1869)	−0.02685 (0.1405)	0.1165 (0.1774)	1.5339 (0.1762)	0.006499 (0.00421)	0.01416	0.9930
WC	***OLS***	9.1108 (3.9704)	0.2334 (0.2062)	1.6043 (0.7449)	0.7174 (0.1613)	−0.3563 (0.3153)	−0.2592 (0.3029)	−0.03416 (0.00629)	0.01085	0.9895
	FGLS	−10.2989 (2.4189)	0.03734 (0.1107)	1.8176 (0.4503)	0.6572 (0.1011)	−0.4358 (0.1912)	0.02904 (0.1828)	−0.02867 (0.00373)	0.008837	0.9881
Pooled	***OLS***	3.1567 (0.1377)	0.08692 (0.01058)	−0.02956 (0.03405)	0.4922 (0.04167)	0.06092 (0.03833)	0.3676 (0.04018)	−0.01746 (0.00304)	0.05558	0.9926
	FGLS	3.1089 (0.0208)	0.08076 (0.005148)	−0.01797 (0.006186)	0.3728 (0.01311)	0.1221 (0.00557)	0.4206 (0.01442)	−0.01506 (0.00101)	NA	0.9882[a]
	FGLS Het.	3.0977 (0.1233)	0.08646 (0.01144)	−0.02141 (0.02830)	0.03874 (0.03529)	0.1215 (0.02805)	0.4032 (0.03410)	−0.01529 (0.00256)	NA	0.9875[a]

[a] R^2 for models fit by FGLS is computed using $1 - 9/\mathrm{tr}(\mathbf{S}^{-1}\mathbf{S}_{yy})$

~~10.2.7~~ 10.3 SEEMINGLY UNRELATED GENERALIZED REGRESSION MODELS

In principle, the SUR model can accommodate heteroscedasticity as well as autocorrelation. Bartels and Fiebig (1992) suggested the generalized SURmodel, $\mathbf{\Omega} = \mathbf{A}[\mathbf{\Sigma} \otimes \mathbf{I}]\mathbf{A}'$ where $\mathbf{A}$ is a block diagonal matrix. Ideally, $\mathbf{A}$ is made a function of measured characteristics of the individual and a separate parameter vector, $\boldsymbol{\theta}$, so that the model can be estimated in stages. In a first step, OLS residuals could be used to form a preliminary estimator of $\boldsymbol{\theta}$, then the data are transformed to homoscedasticity, leaving $\mathbf{\Sigma}$ and $\boldsymbol{\beta}$ to be estimated at subsequent steps using transformed data. One application along these lines is the random parameters model of Fiebig, Bartels, and Aigner (1991); (9-50) shows how the random parameters model induces heteroscedasticity. Another application is Mandy and Martins-Filho (1993), who specified $\sigma_{ij}(t) = \mathbf{z}_{ij}(t)'\boldsymbol{\alpha}_{ij}$. (The linear specification of a variance does present some problems, as a negative value is not precluded.) Kumbhakar and Heshmati (1996) proposed a cost and demand system that combined the translog model of Section 10.4.2 with the complete equation system in 10.4.1. In their application, only the cost equation was specified to include a heteroscedastic disturbance. and/

Autocorrelation in the disturbances of regression models usually arises as a particular feature of the time-series model. It is among the properties of the time series. (We will explore this aspect of the model specification in detail in Chapter 20.) In the multiple equation models examined in this chapter, the time-series properties of the data are usually not the main focus of the investigation. The main advantage of the SUR specification is its treatment of the correlation *across* observations at a particular point in time. Frequently, panel data specifications, such as those in examples 3 and 4 in the Introduction, can also be analyzed in the framework of the SUR model of this chapter. In these cases, there may be persistent effects in the disturbances, but here, again, those effects are often viewed as a consequence of the presence of latent, time invariant heterogeneity. Nonetheless, because the multiple equations models examined in this chapter often do involve moderately long time series, it is appropriate to deal at least somewhat more formally with autocorrelation. Opinions differ on the appropriateness of "corrections" for autocorrelation. At one extreme is Mizon (1995) who argues forcefully that autocorrelation arises as a consequence of a remediable failure to include dynamic effects in the model. However, in a system of equations, the analysis that leads to this conclusion is going to be far more complex than in a single equation model.[16] Suffice to say, the issue remains to be settled conclusively.

chapter/

FN 16

[16] Dynamic SUR models in the spirit of Mizon's admonition were proposed by Anderson and Blundell (1982). A few recent applications are Kiviet, Phillips, and Schipp (1995) and DesChamps (1998). However, relatively little work has been done with dynamic SUR models. The VAR models in Section 21.6 are an important group of applications, but they come from a different analytical framework. Likewise, the panel data applications noted in the Introduction and in Section 9.8.5 would fit into the modeling framework we are developing here. However, in these applications, the regressions are "actually" unrelated—the authors did not model the cross-unit correlation that is the central focus of this chapter. Related results may be found in Guilkey and Schmidt (1973), Guilkey (1974), Berndt and Savin (1977), Moschino and Moro (1994), McLaren (1996), and Holt (1998).

statistic is n times the uncentered R^2 in the regression of $\mathbf{e}_*$ on $\mathbf{X}_*^0$. Many Lagrange multiplier statistics are computed in this fashion.

Example 11.8 Hypothesis Tests in a Nonlinear Regression Model
We test the hypothesis $H_0: \gamma = 1$ in the consumption function of Section 11.3.1.

- ***F* statistic.** The F statistic is

$$F[1, 204-3] = \frac{(1{,}536{,}321.881 - 504{,}403.17)\,/\,1}{504{,}403.17\,/\,(204-3)} = 411.29.$$

The critical value from the tables is 3.84, so the hypothesis is rejected.

- **Wald statistic.** For our example, the Wald statistic is based on the distance of $\hat{\gamma}$ from 1 and is simply the square of the asymptotic t ratio we computed in the example:

$$W = \frac{(1.244827 - 1)^2}{0.01205^2} = 412.805.$$

The critical value from the chi-squared table is 3.84.

- **Lagrange multiplier.** For our example, the elements in $\mathbf{x}_i^*$ are

$$\mathbf{x}_i^* = [1, Y^{\gamma}, \beta\gamma Y^{\gamma} \ln Y].$$

To compute this at the restricted estimates, we use the ordinary least squares estimates for α and β and 1 for γ so that

$$\mathbf{x}_i^* = [1, Y, \beta Y \ln Y].$$

The residuals are the least squares residuals computed from the linear regression. Inserting the values given earlier, we have

$$\text{LM} = \frac{996{,}103.9}{(1{,}536{,}321.881\,/\,204)} = 132.267.$$

As expected, this statistic is also larger than the critical value from the chi-squared table.

~~11.5~~ 10.4 NONLINEAR SYSTEMS OF EQUATIONS

We now consider estimation of nonlinear systems of equations. The underlying theory is essentially the same as that for linear systems. As such, most of the following will describe practical aspects of estimation. Consider estimation of the parameters of the equation system

$$\begin{aligned}
\mathbf{y}_1 &= \mathbf{h}_1(\boldsymbol{\beta}, \mathbf{X}) + \boldsymbol{\varepsilon}_1,\\
\mathbf{y}_2 &= \mathbf{h}_2(\boldsymbol{\beta}, \mathbf{X}) + \boldsymbol{\varepsilon}_2,\\
&\ \ \vdots\\
\mathbf{y}_M &= \mathbf{h}_M(\boldsymbol{\beta}, \mathbf{X}) + \boldsymbol{\varepsilon}_M.
\end{aligned} \qquad (\cancel{11}\,10\text{-}23)$$

[Note the analogy to (10-19).]

There are M equations in total, to be estimated with $t = 1, \ldots, T$ observations. There are K parameters in the model. No assumption is made that each equation has "its own" parameter vector; we simply use some of or all the K elements in $\boldsymbol{\beta}$ in each equation. Likewise, there is a set of T observations on each of P independent variables $\mathbf{x}_p$, $p = 1, \ldots, P$, some of or all that appear in each equation. For convenience, the

10-21

equations are written generically in terms of the full $\boldsymbol{\beta}$ and $\mathbf{X}$. The disturbances are assumed to have zero means and contemporaneous covariance matrix $\boldsymbol{\Sigma}$. We will leave the extension to autocorrelation for more advanced treatments.

In the multivariate regression model, if $\boldsymbol{\Sigma}$ is known, then the generalized least squares estimator of $\boldsymbol{\beta}$ is the vector that minimizes the generalized sum of squares

$$\boldsymbol{\varepsilon}(\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}(\boldsymbol{\beta}) = \sum_{i=1}^{M}\sum_{j=1}^{M} \sigma^{ij}[\mathbf{y}_i - \mathbf{h}_i(\boldsymbol{\beta}, \mathbf{X})]'[\mathbf{y}_j - \mathbf{h}_j(\boldsymbol{\beta}, \mathbf{X})], \qquad (\cancel{11}\ 10\text{-}24)$$

3

where $\boldsymbol{\varepsilon}(\boldsymbol{\beta})$ is an $MT \times 1$ vector of disturbances obtained by stacking the equations, $\boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \mathbf{I}$, and σ^{ij} is the ijth element of $\boldsymbol{\Sigma}^{-1}$. [See (10-7).] As we did in Section 11.2, 7.2.6 define the pseudoregressors as the derivatives of the $\mathbf{h}(\boldsymbol{\beta}, \mathbf{X})$ functions with respect to $\boldsymbol{\beta}$. That is, linearize each of the equations. Then the first-order condition for minimizing this sum of squares is

$$\frac{\partial \boldsymbol{\varepsilon}(\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{M}\sum_{j=1}^{M} \sigma^{ij}[2\mathbf{X}_i^{0\prime}(\boldsymbol{\beta})\boldsymbol{\varepsilon}_j(\boldsymbol{\beta})] = \mathbf{0}, \qquad (\cancel{11}\ 10\text{-}25)$$

where $\mathbf{X}_i^0(\boldsymbol{\beta})$ is the $T \times K$ matrix of pseudoregressors from the linearization of the ith
7.2.6 equation. (See Section 11.2.3.) If any of the parameters in $\boldsymbol{\beta}$ do not appear in the ith equation, then the corresponding column of $\mathbf{X}_i^0(\boldsymbol{\beta})$ will be a column of zeros.

This problem of estimation is doubly complex. In almost any circumstance, solution will require an iteration using one of the methods discussed in Appendix E. Second, of course, is that $\boldsymbol{\Sigma}$ is not known and must be estimated. Remember that efficient estimation in the multivariate regression model does not require an efficient estimator of $\boldsymbol{\Sigma}$, only a consistent one. Therefore, one approach would be to estimate the parameters of each equation separately using nonlinear least squares. This method will be inefficient if any of the equations share parameters, since that information will be ignored. But at this step, consistency is the objective, not efficiency. The resulting residuals can then be used to compute

$$\mathbf{S} = \frac{1}{T}\mathbf{E}'\mathbf{E}. \qquad (\cancel{11}\ 10\text{-}26)$$

The second step of FGLS is the solution of (~~11~~ 10-25), which will require an iterative procedure once again and can be based on $\mathbf{S}$ instead of $\boldsymbol{\Sigma}$. With well-behaved pseudoregressors, this second-step estimator is fully efficient. Once again, the same theory used
FN 17 for FGLS in the linear, single-equation case applies here.[~~10~~ 17] Once the FGLS estimator is obtained, the appropriate asymptotic covariance matrix is estimated with

$$\text{Est. Asy. Var}[\hat{\boldsymbol{\beta}}] = \left[\sum_{i=1}^{M}\sum_{j=1}^{M} s^{ij}\mathbf{X}_i^0(\boldsymbol{\beta})'\mathbf{X}_j^0(\boldsymbol{\beta})\right]^{-1}. \qquad (\cancel{11}\ 10\text{-}27)$$

~~10~~ 17 Neither the nonlinearity nor the multiple equation aspect of this model brings any new statistical issues to the fore. By stacking the equations, we see that this model is simply a variant of the nonlinear regression model with the added complication of a nonscalar disturbance covariance matrix, which we analyzed in Chapter 8. The new complications are primarily practical.

10-22

There is a possible flaw in the strategy just outlined. It may not be possible to fit all the equations individually by nonlinear least squares. It is conceivable that identification of some of the parameters requires joint estimation of more than one equation. But as long as the full system identifies all parameters, there is a simple way out of this problem. Recall that all we need for our first step is a consistent set of estimators of the elements of $\boldsymbol{\beta}$. It is easy to show that the preceding defines a GMM estimator (see Chapter ~~15~~.) We can use this result to devise an alternative, simple strategy. The weighting of the sums of squares and cross products in (~~11~~-24) by σ^{ij} produces an efficient estimator of $\boldsymbol{\beta}$. Any other weighting based on some positive definite $\mathbf{A}$ would produce consistent, although inefficient, estimates. At this step, though, efficiency is secondary, so the choice of $\mathbf{A} = \mathbf{I}$ is a convenient candidate. Thus, for our first step, we can find $\boldsymbol{\beta}$ to minimize 13

10

$$\boldsymbol{\varepsilon}(\boldsymbol{\beta})'\boldsymbol{\varepsilon}(\boldsymbol{\beta}) = \sum_{i=1}^{M}[\mathbf{y}_i - \mathbf{h}_i(\boldsymbol{\beta}, \mathbf{X})]'[\mathbf{y}_i - \mathbf{h}_i(\boldsymbol{\beta}, \mathbf{X})] = \sum_{i=1}^{M}\sum_{t=1}^{T}[y_{it} - h_i(\boldsymbol{\beta}, \mathbf{x}_{it})]^2.$$

(This estimator is just pooled nonlinear least squares, where the regression function varies across the sets of observations.) This step will produce the $\hat{\boldsymbol{\beta}}$ we need to compute $\mathbf{S}$.

11.6 TWO-STEP NONLINEAR LEAST SQUARES ESTIMATION

In this section, we consider the case in which the nonlinear regression model depends on a second set of parameters that is estimated separately.

The model is

$$y = h(\mathbf{x}, \boldsymbol{\beta}, \mathbf{w}, \boldsymbol{\gamma}) + \varepsilon.$$

We consider cases in which the auxiliary parameter $\boldsymbol{\gamma}$ is estimated separately in a model that depends on an additional set of variables $\mathbf{w}$. This first step might be a least squares regression, a nonlinear regression, or a maximum likelihood estimation. The parameters $\boldsymbol{\gamma}$ will usually enter $h(.)$ through some function of $\boldsymbol{\gamma}$ and $\mathbf{w}$, such as an expectation. The second step then consists of a nonlinear regression of y on $h(\mathbf{x}, \boldsymbol{\beta}, \mathbf{w}, \mathbf{c})$ in which $\mathbf{c}$ is the first-round estimate of $\boldsymbol{\gamma}$. To put this in context, we will develop an example.

The estimation procedure is as follows.

1. Estimate $\boldsymbol{\gamma}$ by least squares, nonlinear least squares, or maximum likelihood. We assume that this estimator, however obtained, denoted $\mathbf{c}$, is consistent and asymptotically normally distributed with asymptotic covariance matrix $\frac{1}{n}\mathbf{V}_c$. Let $\hat{\mathbf{V}}_c$ be any appropriate estimator of $\mathbf{V}_c$.
2. Estimate $\boldsymbol{\beta}$ by nonlinear least squares regression of y on $h(\mathbf{x}, \boldsymbol{\beta}, \mathbf{w}, \mathbf{c})$. Let $\frac{\sigma^2}{n}\mathbf{V}_b$ be the asymptotic covariance matrix of this estimator of $\boldsymbol{\beta}$, assuming $\boldsymbol{\gamma}$ is known and let $\frac{s^2}{n}\hat{\mathbf{V}}_b$ be any appropriate estimator of $\frac{\sigma^2}{n}\mathbf{V}_b = \frac{\sigma^2}{n}\left(\frac{\mathbf{X}^{0\prime}\mathbf{X}^0}{n}\right)^{-1}$, where $\mathbf{X}^0$ is the matrix of pseudoregressors evaluated at the true parameter values $\mathbf{x}_i^0 = \partial h(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{w}_i, \boldsymbol{\gamma})/\partial\boldsymbol{\beta}$.

10-23

TABLE 10.2 Coefficient Estimates in SUR Model for Hospital Costs

	Coefficient on Variable in the Equation				
Equation	*DIS87*	*DIS88*	*DIS89*	*DIS90*	*DIS91*
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.)

10.5 ~~10.4~~ SYSTEMS OF DEMAND EQUATIONS: SINGULAR SYSTEMS

FN 18

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.

KT

Aus KT "systems of demand equations" not in chap. list

FN 19

Example 10.2 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 multi*variate* 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).

Other

10-24

FN 20

FN 21

theory.[20] The numerous constraints in the system of demand equations presented earlier give an example. A second intrinsic feature of many of these models is that the disturbance covariance matrix $\mathbf{\Sigma}$ is singular.[21]

10.~~4~~5.1 COBB–DOUGLAS COST FUNCTION ~~(EXAMPLE 6.3 CONTINUED)~~

Consider a **Cobb–Douglas** production function,

$$Q = \alpha_0 \prod_{i=1}^{M} x_i^{\alpha_i}.$$

Profit maximization with an exogenously determined output price calls for the firm to maximize output for a given cost level C (or minimize costs for a given output Q). The Lagrangean for the maximization problem is

$$\Lambda = \alpha_0 \prod_{i=1}^{M} x_i^{\alpha_i} + \lambda(C - \mathbf{p}'\mathbf{x}),$$

where $\mathbf{p}$ is the vector of M factor prices. The necessary conditions for maximizing this function are

$$\frac{\partial \Lambda}{\partial x_i} = \frac{\alpha_i Q}{x_i} - \lambda p_i = 0 \quad \text{and} \quad \frac{\partial \Lambda}{\partial \lambda} = C - \mathbf{p}'\mathbf{x} = 0.$$

The joint solution provides $x_i(Q, \mathbf{p})$ and $\lambda(Q, \mathbf{p})$. The total cost of production is

$$\sum_{i=1}^{M} p_i x_i = \sum_{i=1}^{M} \frac{\alpha_i Q}{\lambda}.$$

The cost share allocated to the ith factor is

$$\frac{p_i x_i}{\sum_{i=1}^{M} p_i x_i} = \frac{\alpha_i}{\sum_{i=1}^{M} \alpha_i} = \beta_i. \tag{10-28}$$

FN 22

The full model is[22]

$$\begin{aligned} \ln C &= \beta_0 + \beta_q \ln Q + \sum_{i=1}^{M} \beta_i \ln p_i + \varepsilon_c, \\ s_i &= \beta_i + \varepsilon_i, \quad i = 1, \ldots, M. \end{aligned} \tag{10-29}$$

---

[20] This inconsistency does not imply that the theoretical restrictions are not testable or that the unrestricted model cannot be estimated. Sometimes, the meaning of the model is ambiguous without the restrictions, however. Statistically rejecting the restrictions implied by the theory, which were used to derive the econometric model in the first place, can put us in a rather uncomfortable position. For example, in a study of utility functions, Christensen, Jorgenson, and Lau (1975), after rejecting the cross-equation symmetry of a set of commodity demands, stated, "With this conclusion we can terminate the test sequence, since these results invalidate the theory of demand" (p. 380). See Silver and Ali (1989) for discussion of testing symmetry restrictions. The theory and the model may also conflict in other ways. For example, Stone's loglinear expenditure system in Example 10.7 does not conform to any theoretically valid utility function. See Goldberger (1987).

[21] Denton (1978) examines several of these cases.

[22] We leave as an exercise the derivation of β_0, which is a mixture of all the parameters, and β_q, which equals $1/\Sigma_m \alpha_m$.

10-25

6.6

By construction, $\sum_{i=1}^{M} \beta_i = 1$ and $\sum_{i=1}^{M} s_i = 1$. (This is the cost function analysis begun in Example 6.3. We will return to that application below.) The cost shares will also sum identically to one in the data. It therefore follows that $\sum_{i=1}^{M} \varepsilon_i = 0$ at every data point, so the system is singular. For the moment, ignore the cost function. Let the $M \times 1$ disturbance vector from the shares be $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M]'$. Because $\boldsymbol{\varepsilon}'\mathbf{i} = 0$, where $\mathbf{i}$ is a column of 1s, it follows that $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{i}] = \boldsymbol{\Sigma}\mathbf{i} = \mathbf{0}$, which implies that $\boldsymbol{\Sigma}$ is singular. Therefore, the methods of the previous sections cannot be used here. (You should verify that the *sample* covariance matrix of the OLS residuals will also be singular.)

The solution to the singularity problem appears to be to drop one of the equations, estimate the remainder, and solve for the last parameter from the other $M - 1$. The constraint $\sum_{i=1}^{M} \beta_i = 1$ states that the cost function must be homogeneous of degree one in the prices, a theoretical necessity. If we impose the constraint

$$\beta_M = 1 - \beta_1 - \beta_2 - \cdots - \beta_{M-1}, \tag{10-30}$$

then the system is reduced to a nonsingular one:

$$\ln\left(\frac{C}{p_M}\right) = \beta_0 + \beta_q \ln Q + \sum_{i=1}^{M-1} \beta_i \ln\left(\frac{p_i}{p_M}\right) + \varepsilon_c,$$

$$s_i = \beta_i + \varepsilon_i, \quad i = 1, \ldots, M-1.$$

Ao: KT "invariant" not in ch. list

This system provides estimates of β_0, β_q, and $\beta_1, \ldots, \beta_{M-1}$. The last parameter is estimated using (10-30). It is immaterial which factor is chosen as the numeraire. Both FGLS and **maximum likelihood,** which can be obtained by iterating FGLS or by direct maximum likelihood estimation, are **invariant** to which factor is chosen as the numeraire.[23]

KT FN 23

Ao: Should Example 6.3 be 6.6?

6.6

Nerlove's (1963) study of the electric power industry that we examined in Example 6.3 provides an application of the Cobb–Douglas cost function model. His ordinary least squares estimates of the parameters were listed in Example 6.3. Among the results are (unfortunately) a negative capital coefficient in three of the six regressions. Nerlove also found that the simple Cobb–Douglas model did not adequately account for the relationship between output and average cost. Christensen and Greene (1976) further analyzed the Nerlove data and augmented the data set with cost share data to estimate the complete **demand system.** Appendix Table F10.1 lists Nerlove's 145 observations with Christensen and Greene's cost share data. Cost is the total cost of generation in millions of dollars, output is in millions of kilowatt-hours, the capital price is an index of construction costs, the wage rate is in dollars per hour for production and maintenance, the fuel price is an index of the cost per Btu of fuel purchased by the firms, and the data reflect the 1955 costs of production. The regression estimates are given in Table 10.~~3~~2.

KT F6.2 TB 10.2

Least squares estimates of the Cobb–Douglas cost function are given in the first column.[24] The coefficient on capital is negative. Because $\beta_i = \beta_q \partial \ln Q / \partial \ln x_i$—that is, a positive multiple of the output elasticity of the ith factor—this finding is troubling.

FN 24

[23]The invariance result is proved in Barten (1969). Some additional results on the method are given by Revankar (1976), Deaton (1986), Powell (1969), and McGuire et al. (1968).

[24]Results based on Nerlove's full data set are given in Example 6.3. ~~We have recomputed the values given in Table 6.3. Note that Nerlove used base 10 logs while we have used natural logs in our computations.~~

6.6

10-26

TABLE 10.~~1~~2 Regression Estimates (standard errors in parentheses)

	Ordinary Least Squares				*Multivariate Regression*			
β_0	−4.686	(0.885)	−3.764	(0.702)	−7.069	(0.107)	−5.707	(0.165)
β_q	0.721	(0.0174)	0.153	(0.0618)	0.766	(0.0154)	0.238	(0.0587)
β_{qq}	—		0.0505	(0.00536)	—		0.0451	(0.00508)
β_k	−0.00847	(0.191)	0.0739	(0.150)	0.424	(0.00946)	0.424	(0.00944)
β_l	0.594	(0.205)	0.481	(0.161)	0.106	(0.00386)	0.106	(0.00382)
β_f	0.414	(0.0989)	0.445	(0.0777)	0.470	(0.0101)	0.470	(0.0100)
R^2	0.9316		0.9581		—	—		

The third column presents the constrained FGLS estimates. To obtain the constrained estimator, we set up the model in the form of the pooled SUR estimator in (10-19);

$$\mathbf{y} = \begin{bmatrix} \ln(C/P_f) \\ \mathbf{s}_k \\ \mathbf{s}_l \end{bmatrix} = \begin{bmatrix} \mathbf{i} & \ln \mathbf{Q} & \ln(\mathbf{P}_k/\mathbf{P}_f) & \ln(\mathbf{P}_l/\mathbf{P}_f) \\ \mathbf{0} & \mathbf{0} & \mathbf{i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{i} \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_q \\ \beta_k \\ \beta_l \end{pmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_c \\ \boldsymbol{\varepsilon}_k \\ \boldsymbol{\varepsilon}_l \end{bmatrix}$$

[There are 3(145) = 435 observations in the data matrices.] The estimator is then FGLS as shown in (10-21). An additional column is added for the log quadratic model. Two things to note are the dramatically smaller standard errors and the now positive (and reasonable) estimate of the capital coefficient. The estimates of economies of scale in the basic Cobb–Douglas model are $1/\beta_q = 1.39$ (column 1) and 1.31 (column 3), which suggest some increasing returns to scale. Nerlove, however, had found evidence that at extremely large firm sizes, economies of scale diminished and eventually disappeared. To account for this (essentially a classical U-shaped average cost curve), he appended a quadratic term in log output in the cost function. The single equation and multivariate regression estimates are given in the second and fourth sets of results.

The quadratic output term gives the cost function the expected U-shape. We can determine the point where average cost reaches its minimum by equating $\partial \ln C/\partial \ln Q$ to 1. This is $Q^* = \exp[(1 - \beta_q)/(2\beta_{qq})]$. For the multivariate regression, this value is $Q^* = 4665$. About 85 percent of the firms in the sample had output less than this, so by these estimates, most firms in the sample had not yet exhausted the available economies of scale. Figure 10.1 shows predicted and actual average costs for the sample. (To obtain a reasonable scale, the smallest one third of the firms are omitted from the figure.) Predicted average costs are computed at the sample averages of the input prices. The figure does reveal that that beyond a quite small scale, the economies of scale, while perhaps statistically significant, are economically quite small.

FIG 10.1

10.~~4~~5.2 FLEXIBLE FUNCTIONAL FORMS: THE TRANSLOG COST FUNCTION

The literatures on production and cost and on utility and demand have evolved in several directions. In the area of models of producer behavior, the classic paper by Arrow et al. (1961) called into question the inherent restriction of the Cobb–Douglas model that all elasticities of factor substitution are equal to 1. Researchers have since developed

10-27

replace with figure on next page
ms p 10-28

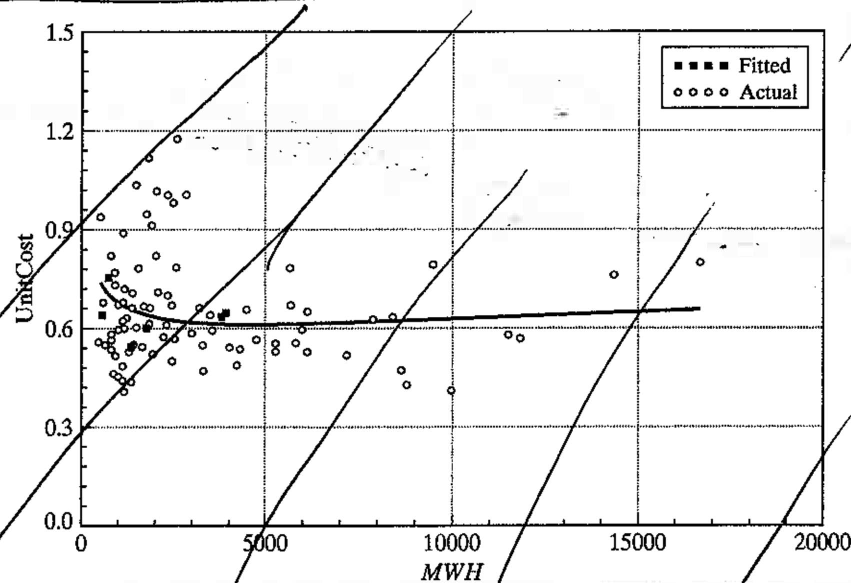

FIGURE 10.1 Predicted and Actual Average Costs.

AQ: KT "flexible functions" not in Ch. list

KT

FN 25

FN 26

numerous **flexible functions** that allow substitution to be unrestricted (i.e., not even constant).[25] Similar strands of literature have appeared in the analysis of commodity demands.[26] In this section, we examine in detail a model of production.

Suppose that production is characterized by a production function, $Q = f(\mathbf{x})$. The solution to the problem of minimizing the cost of producing a specified output rate given a set of factor prices produces the cost-minimizing set of factor demands $x_i = x_i(Q, \mathbf{p})$. The total cost of production is given by the cost function,

$$C = \sum_{i=1}^{M} p_i x_i(Q, \mathbf{p}) = C(Q, \mathbf{p}). \tag{10-31}$$

KT

If there are **constant returns to scale,** then it can be shown that $C = Qc(\mathbf{p})$ or

$$C/Q = c(\mathbf{p}),$$

FN 27

where $c(\mathbf{p})$ is the unit or average cost function.[27] The cost-minimizing factor demands are obtained by applying **Shephard's** (1970) **lemma,** which states that if $C(Q, \mathbf{p})$ gives the minimum total cost of production, then the cost-minimizing set of factor demands

KT

[25]See, in particular, Berndt and Christensen (1973). Two useful surveys of the topic are Jorgenson (1983) and Diewert (1974).

[26]See, for example, Christensen, Jorgenson, and Lau (1975) and two surveys, Deaton and Muellbauer (1980) and Deaton (1983). Berndt (1990) contains many useful results.

[27]The Cobb–Douglas function of the previous section gives an illustration. The restriction of constant returns to scale is $\beta_q = 1$, which is equivalent to $C = Qc(\mathbf{p})$. Nerlove's more general version of the cost function allows nonconstant returns to scale. See Christensen and Greene (1976) and Diewert (1974) for some of the formalities of the cost function and its relationship to the structure of production.

10-28

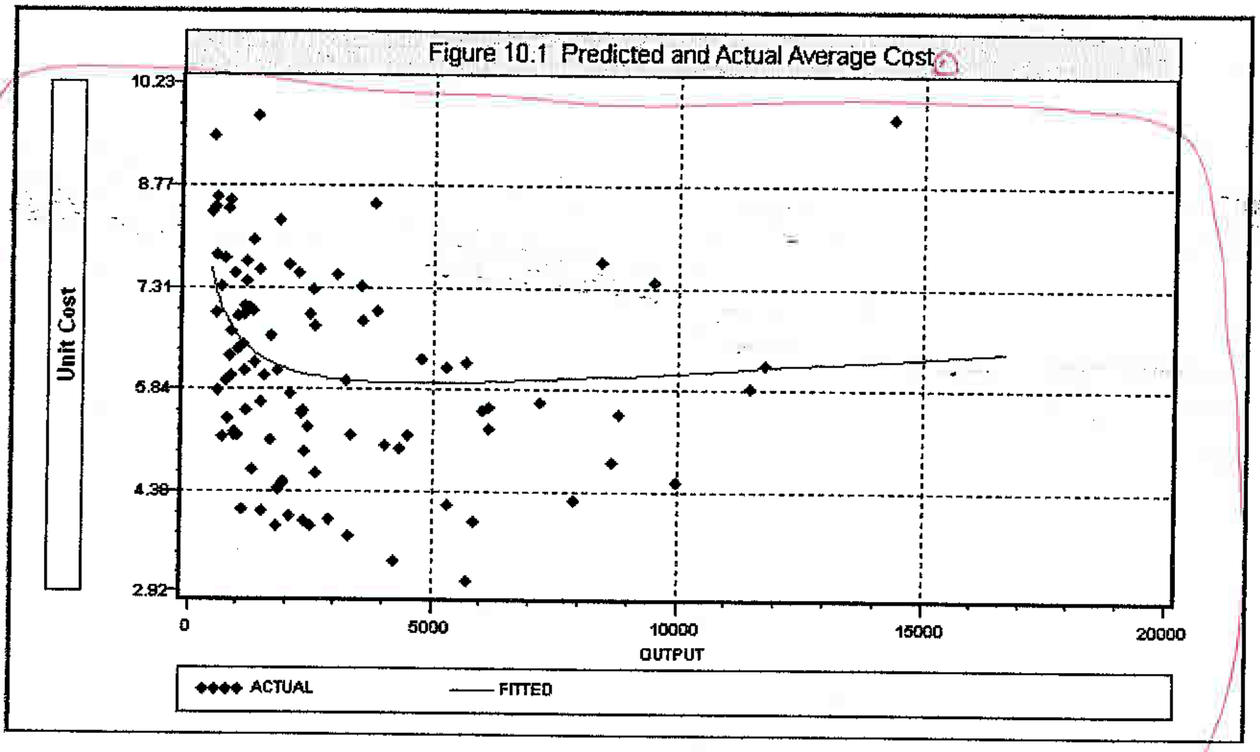

Figure 10.1 Predicted and Actual Average Cost
Unit Cost
10.23
8.77
7.31
5.84
4.38
2.92
0
5000
10000
15000
20000
OUTPUT
ACTUAL
FITTED

10-29

is given by

$$x_i^* = \frac{\partial C(Q, \mathbf{p})}{\partial p_i} = \frac{Q \partial c(\mathbf{p})}{\partial p_i}. \tag{10-32}$$

Alternatively, by differentiating logarithmically, we obtain the cost-minimizing factor cost shares:

$$s_i = \frac{\partial \ln C(Q, \mathbf{p})}{\partial \ln p_i} = \frac{p_i x_i}{C}. \tag{10-33}$$

With constant returns to scale, $\ln C(Q, \mathbf{p}) = \ln Q + \ln c(\mathbf{p})$, so

$$s_i = \frac{\partial \ln c(\mathbf{p})}{\partial \ln p_i}. \tag{10-34}$$

In many empirical studies, the objects of estimation are the elasticities of factor substitution and the own price elasticities of demand, which are given by

$$\theta_{ij} = \frac{c(\partial^2 c/\partial p_i \partial p_j)}{(\partial c/\partial p_i)(\partial c/\partial p_j)}$$

and

$$\eta_{ii} = s_i \theta_{ii}.$$

By suitably parameterizing the cost function (10-31) and the cost shares (10-34), we obtain an M or $M+1$ equation econometric model that can be used to estimate these quantities.[28]

FN 28

The transcendental logarithmic, or **translog function** is the most frequently used flexible function in empirical work.[29] By expanding $\ln c(\mathbf{p})$ in a second-order **Taylor series** about the point $\ln \mathbf{p} = \mathbf{0}$, we obtain

KT

FN 29

AO: KTS "translog function" & "Taylor Series" not in ch. list

$$\ln c \approx \beta_0 + \sum_{i=1}^{M} \left(\frac{\partial \ln c}{\partial \ln p_i}\right) \log p_i + \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \left(\frac{\partial^2 \ln c}{\partial \ln p_i \, \partial \ln p_j}\right) \ln p_i \ln p_j, \tag{10-35}$$

where all derivatives are evaluated at the expansion point. If we treat these derivatives as the coefficients, then the cost function becomes

$$\begin{aligned} \ln c = \beta_0 &+ \beta_1 \ln p_1 + \cdots + \beta_M \ln p_M + \delta_{11}\left(\tfrac{1}{2}\ln^2 p_1\right) + \delta_{12} \ln p_1 \ln p_2 \\ &+ \delta_{22}\left(\tfrac{1}{2}\ln^2 p_2\right) + \cdots + \delta_{MM}\left(\tfrac{1}{2}\ln^2 p_M\right). \end{aligned} \tag{10-36}$$

[28] The cost function is only one of several approaches to this study. See Jorgenson (1983) for a discussion.

2.4 [29] See Example 2.4. The function was developed by Kmenta (1967) as a means of approximating the CES production function and was introduced formally in a series of papers by Berndt, Christensen, Jorgenson, and Lau, including Berndt and Christensen (1973) and Christensen et al. (1975). The literature has produced something of a competition in the development of exotic functional forms. The translog function has remained the most popular, however, and by one account, Guilkey, Lovell, and Sickles (1983) is the most reliable of several available alternatives. See also Example 5.2.

5.4