10

SYSTEMS OF REGRESSION 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, we will want to consider the several models jointly. Here are example:

1. **Set of Regression Equations.** Munnell's (1990) model for output by the 48 contiguous states in the U.S., m, at time t is

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

   where the variables are labor and public capital. 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 (t,s) within a state. It would be natural as well for observations at a point in time to be correlated across states (m,n), at least for some states. An important question is whether it is valid to assume that the coefficient vector is the same for all states in the sample.

2. **Identical Regressors.** 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 provides some information about the excess return of security j, at least for some j's. It may be useful to estimate the equations jointly rather than ignore this connection. The fact that the right-hand side, $[constant, r_{mt} - r_{ft}]$, is the same for all i makes this model an interesting special case of the more general set of regressions.

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

$$y_{mt} = \alpha_m + \mathbf{x}_{mt}'\boldsymbol{\beta}_m + \gamma_m y_{m,t-1} + \varepsilon_{mt}.$$

   Nair-Reichert and Weinhold's (2001) cross-country analysis of growth in 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. **System of Demand Equations.** In a model of production, the optimization conditions of economic theory imply that, if a firm faces a set of factor prices **p**, then its set of cost-minimizing factor demands for producing output Q will be a set of M equations of the form $x_m = f_m(Q, \mathbf{p})$. The empirical model is

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

where $\boldsymbol{\theta}$ is a vector of parameters that are part of the technology and ε_m 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 has cross-equation restrictions. Estimating the equations separately will waste the information that the same set of parameters appears in all the equations.

5. **Vector Autoregression.** A useful formulation that appears in many macroeconomics applications is the vector autoregression, or VAR. In Chapter 13, we will examine a model of Swedish municipal government fiscal activities in the form

$$\begin{aligned} S_{m,t} &= \alpha_1 + \gamma_{11} S_{m,t-1} + \gamma_{12} R_{m,t-1} + \gamma_{13} G_{m,t-1} + \varepsilon_{S,m,t}, \\ R_{m,t} &= \alpha_2 + \gamma_{21} S_{m,t-1} + \gamma_{22} R_{m,t-1} + \gamma_{23} G_{m,t-1} + \varepsilon_{R,m,t}, \\ G_{m,t} &= \alpha_3 + \gamma_{31} S_{m,t-1} + \gamma_{32} R_{m,t-1} + \gamma_{33} G_{m,t-1} + \varepsilon_{G,m,t}, \end{aligned}$$

where S, R, and G are spending, tax revenues, and grants, respectively, for municipalities m in period t. VARs without restrictions are similar to Example 2 above. The dynamic equations can be used to trace the influences of shocks in a system as they exert their influence through time.

6. **Linear panel data model.** In Chapter 11, we will examine models for **panel data** – $t = 1, \ldots, T$ repeated observations on individuals m, of the form

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

In Example 11.1, we consider a wage equation,

$$\ln Wage_{mt} \quad = \quad \beta_1 + \beta_2\, Experience_{mt} + \ \ldots\ + \mathbf{x}_{mt}'\boldsymbol{\beta} + \varepsilon_{mt}.$$

For some purposes, it is useful to consider this model as a set of T regression equations, one for each period. Specification of the model focuses on correlations of the unobservables in ε_{mt} across periods and with dynamic behavior of $\ln Wage_{mt}$.

7. **Simultaneous Equations System.** A common form of a model for equilibrium in a market would be

$$\begin{aligned} Q_{Demand} &= \alpha_1 + \alpha_2\, Price + \alpha_3\, Income + \mathbf{d}'\boldsymbol{\alpha} + \varepsilon_{Demand}, \\ Q_{Supply} &= \beta_1 + \beta_2\, Price + \beta_3\, FactorPrice + \mathbf{s}'\boldsymbol{\beta} + \varepsilon_{Supply}, \\ Q_{Equilibrium} &= Q_{Demand} = Q_{Supply}, \end{aligned}$$

where **d** and **s** are exogenous 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 into the endogenous variables framework developed in Chapter 8. The multiple equations framework developed in this chapter provides additional results for estimating "simultaneous equations models" such as this.

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 set of parameters and examines efficient estimation techniques and the special case in which the coefficients are the same in all equations. Production and consumer demand models are special cases of the general model in which the equations obey an *adding-up constraint* that has implications for specification and estimation. Such demand systems are examined in Section 10.3. This section examines an application of the seemingly unrelated regressions model that illustrates the interesting features of empirical demand studies. The seemingly unrelated regressions model is also extended to the translog specification, which forms the platform for many microeconomic studies of production and cost. Finally, Section 10.4 combines the results of Chapter 8 on models with endogenous variables with the development in this chapter of multiple equation systems. In this section, we will develop **simultaneous equations models**. The supply and demand model suggested in Example 6 above, 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 suggested in the Introduction have a common structure, which we may write as

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1, \\ \mathbf{y}_2 &= \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2, \\ &\cdots \\ \mathbf{y}_M &= \mathbf{X}_M\boldsymbol{\beta}_M + \boldsymbol{\varepsilon}_M. \end{aligned}$$

There are M equations and T observations in the sample.[1] The **seemingly unrelated regressions (SUR)** model is

$$\mathbf{y}_m = \mathbf{X}_m\boldsymbol{\beta}_m + \boldsymbol{\varepsilon}_m, \quad m = 1, \dots, M. \tag{10-1}$$

The equations are labeled "seemingly unrelated" because they are linked by the possible correlation of the unobserved disturbances, ε_{mt} and ε_{nt}.[2] By stacking the sets of observations, we obtain

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

[2]See Zellner (1962) who coined the term.

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{X}_M \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \qquad \textbf{(10-2)}$$

The $MT \times 1$ vector of disturbances is

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

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 and nonautocorrelation within each equation,

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

The strict exogeneity assumption is a bit stronger than necessary for present purposes. We could allow more generality by assuming only $E[\boldsymbol{\varepsilon}_m \mid \mathbf{X}_m] = \mathbf{0}$—that is, allowing the disturbances in equation n to be correlated with the regressors in equation m but not equation n. But that extension would not arise naturally in an application. A total of T observations are to be used in estimating the parameters of the M equations. Each equation involves K_m regressors, for a total of $K = \sum_{m=1}^{M} K_m$ in (10-2). We will require $T > K_m$ (so that, if desired, we could fit each equation separately). The data are assumed to be well behaved,as described in Section 4.4.1, so we shall not treat the issue separately here. For the present, we also assume that disturbances are not correlated across periods (or individuals) but may be correlated across equations (at a point in time or for a given individual). Therefore,

$$E[\varepsilon_{mt}\varepsilon_{ns} \mid \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M] = \sigma_{mn}, \quad \text{if } t = s \text{ and } 0 \text{ if } t \neq s.$$

The disturbance formulation for the entire model is

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

$$= \boldsymbol{\Sigma} \otimes \mathbf{I},$$

where

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

is the $M \times M$ covariance matrix of the disturbances for the tth observation, $\boldsymbol{\varepsilon}_t$.

The SUR model thus far assumes that each equation obeys the assumptions of the linear model of Chapter 2—no heteroscedasticity or autocorrelation (within or across equations). Bartels and Fiebig (1992), Bartels and Aigner (1991), Mandy and Martins-Filho (1993), and Kumbhakar (1996) suggested extensions that involved heteroscedasticity within each equation. Autocorrelation of the disturbances of regression models is usually not the focus of the investigation, though Munnell's application to aggregate statewide data might be a natural application.[3] (It might also be a natural candidate for the "spatial autoregression" model of Section 11.7.) All of these extensions are left for more advanced treatments and specific applications.

10.2.1 ORDINARY LEAST SQUARES AND ROBUST INFERENCE

For purposes of developing effective estimation methods, there are two ways to visualize the arrangement of the data. Consider the model in Example 10.2, which examines a cost function for electricity generation. The three equations are

$$\begin{aligned} \ln(C/P_f) &= \alpha_1 + \alpha_2 \ln Q + \alpha_3 \ln(P_k/P_f) + \alpha_4 \ln(P_l/P_f) + \varepsilon_c, \\ s_k &= \beta_1 + \varepsilon_k, \\ s_l &= \gamma_1 + \varepsilon_l, \end{aligned}$$

where C is total cost, P_k, P_l, and P_f are unit prices for capital, labor, and fuel, Q is output, and s_k and s_l are cost shares for capital and labor. (The fourth equation, for s_f, is obtained from $s_k + s_l + s_f = 1$.) There are $T = 145$ observations for each of the $M = 3$ equations. The data may be *stacked by equations* as in the following,

$$\begin{bmatrix} \ln(\mathbf{C}/\mathbf{P_f}) \\ \mathbf{s}_k \\ \mathbf{s}_l \end{bmatrix} = \begin{bmatrix} \mathbf{i} & \ln \mathbf{Q} & \ln(\mathbf{P}_k/\mathbf{P}_f) & \ln(\mathbf{P}_l/\mathbf{P}_f) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{i} \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \beta_1 \\ \gamma_1 \end{pmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_c \\ \boldsymbol{\varepsilon}_k \\ \boldsymbol{\varepsilon}_l \end{bmatrix}, \tag{10-4}$$

$$\begin{bmatrix} \mathbf{y}_c \\ \mathbf{y}_k \\ \mathbf{y}_l \end{bmatrix} = \begin{bmatrix} \mathbf{X}_c & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_l \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta}_c \\ \boldsymbol{\beta}_k \\ \boldsymbol{\beta}_l \end{pmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_c \\ \boldsymbol{\varepsilon}_k \\ \boldsymbol{\varepsilon}_l \end{bmatrix}.$$

Each block of data in the bracketed matrices contains the T observations for equation m. The covariance matrix for the $MT \times 1$ vector of disturbances appears in (10-3). The data may instead be *stacked by observations* by reordering the rows to obtain

$$\mathbf{y} = \begin{bmatrix} \begin{pmatrix} \ln(C/P_f) \\ s_k \\ s_l \end{pmatrix} i = \mathit{firm}\ 1 \\ \cdots \\ \begin{pmatrix} \ln(C/P_f) \\ s_k \\ s_l \end{pmatrix} i = \mathit{firm}\ T \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \begin{pmatrix} \varepsilon_c \\ \varepsilon_k \\ s_l \end{pmatrix} i = \mathit{firm}\ 1 \\ \cdots \\ \begin{pmatrix} \varepsilon_c \\ \varepsilon_k \\ \varepsilon_l \end{pmatrix} i = \mathit{firm}\ T \end{bmatrix} \text{ and } \mathbf{X} \text{ likewise.} \tag{10-5}$$

[3]Dynamic SUR models are proposed by Anderson, and Blundell (1982). Other applications are examined in Kiviet, Phillips, and Schipp (1995), DesChamps (1998), and Wooldridge (2010, p. 194). The VAR models are an important group of applications, but they come from a different analytical framework. 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).

By this arrangement,

$$E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}] = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma} \end{bmatrix} = \mathbf{I} \otimes \boldsymbol{\Sigma}. \tag{10-6}$$

The arrangement in (10-4) will be more convenient for formulating the applications, as in Example 10.4. The format in (10-5) will be more convenient for formulating the estimator and examining its properties.

From (10-2), we can see that with no restrictions on $\boldsymbol{\beta}$, ordinary least squares estimation of $\boldsymbol{\beta}$ will be equation by equation OLS,

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \Rightarrow \mathbf{b}_m = (\mathbf{X}_m'\mathbf{X}_m)^{-1}\mathbf{X}_m'\mathbf{y}_m.$$

Therefore,

$$\mathbf{b}_{\mathrm{m}} = \boldsymbol{\beta}_m + (\mathbf{X}_m'\mathbf{X}_m)^{-1}\mathbf{X}_m'\boldsymbol{\varepsilon}_m.$$

Because this is a simple regression model with homoscedastic and nonautocorrelated disturbances, the familiar estimator of the asymptotic covariance matrix for $(\mathbf{b}_m, \mathbf{b}_n)$ is

$$\hat{\mathbf{V}} = \text{Est.Asy.Cov}[\mathbf{b}_m, \mathbf{b}_n] = s_{mn}(\mathbf{X}_m'\mathbf{X}_m)^{-1}\mathbf{X}_m'\mathbf{X}_n\,(\mathbf{X}_n'\mathbf{X}_n)^{-1}, \tag{10-7}$$

where $s_{mn} = \mathbf{e}_m'\mathbf{e}_n/T$. There is a small ambiguity about the degrees of freedom in s_{mn}. For the diagonal elements, $(T - K_m)$ would be appropriate. One suggestion for the off-diagonal elements that seems natural, but does not produce an unbiased estimator, is $[(T - K_m)(T - K_n)]^{1/2}$.[4]

For inference purposes, Equation (10-7) relies on the two assumptions of homoscedasticity and nonautocorrelation. We can see in (10-6) what features are accommodated and what are not. The estimator does allow a form of heteroscedasticity across equations, in that $\sigma_{mm} \neq \sigma_{nn}$ when $m \neq n$. This is not a real generality, however. For example, in the cost-share equation, it allows the variance of the cost disturbance to be different from the share disturbance, but that would be expected. It does assume that observations are homoscedastic within each equation, in that $\text{E}[\boldsymbol{\varepsilon}_m\boldsymbol{\varepsilon}_m'|\mathbf{X}] = \sigma_{mm}\mathbf{I}$. It allows observations to be correlated across equations, in that $\sigma_{mn} \neq 0$, but it does not allow observations at different times (or different firms in our example) to be correlated. So, the estimator thus far is not generally robust. Robustness to autocorrelation would be the case of lesser interest, save for the panel data models considered in the next chapter. An extension to more general heteroscedasticity might be attractive. We can allow the diagonal matrices in (10-6) to vary arbitrarily or to depend on $\mathbf{X}_m$. The common $\boldsymbol{\Sigma}$ in (10-6) would be replaced with $\boldsymbol{\Sigma}_m$. The estimator in (10-7) would be replaced by

$$\hat{\mathbf{V}}_{Robust} = \text{Est.Asy.Var}[\mathbf{b}] = \left(\sum\nolimits_{t=1}^{T}\mathbf{X}_t'\mathbf{X}_t\right)^{-1}\left(\sum\nolimits_{t=1}^{T}(\mathbf{X}_t'\mathbf{e}_t)(\mathbf{e}_t'\mathbf{X}_t)\right)\left(\sum\nolimits_{t=1}^{T}\mathbf{X}_t'\mathbf{X}_t\right)^{-1}. \tag{10-8}$$

[4]See Srivastava and Giles (1987).

Note $\mathbf{X}_t$ is M rows and $\Sigma_{m=1}^{M} K_m$ columns corresponding to the tth observations for all M equations, while $\mathbf{e}_t$ is an $M \times 1$ vector of OLS residuals based on (10-5). For example, in (10-5), $\mathbf{X}_1$ is the 3×6 matrix,

$$\mathbf{X}_1 = \begin{bmatrix} 1 & \ln Q & \ln(P_k/P_f) & \ln(P_l/P_f) & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{firm\ 1}.$$

Then, (10-8) would be a multiple equation version of the White estimator for arbitrary heteroscedasticity shown in Section 9.4.4.

For testing hypotheses, either within or across equations, of the form H_0: $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$, we can use the Wald statistic,

$$W = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\mathbf{R}\hat{\mathbf{V}}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}),$$

which has a limiting chi-squared distribution with degrees of freedom equal to the number of restrictions. For simple hypotheses involving one coefficient, such as H_0: $\beta_k = 0$, we would report the square root of W as the "asymptotic t ratio," $z_k = \hat{\beta}_k/\text{Asy.S.E.}(\hat{\beta}_k)$ where the asymptotic standard error would be the square root of the diagonal element of $\hat{\mathbf{V}}$. This would have a standard normal distribution in large samples under the null hypothesis.

10.2.2 GENERALIZED LEAST SQUARES

Each equation is, by itself, a linear 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 in (10-2). In (10-3), where the $\mathbf{I}$ matrix is $T \times T$, the $MT \times MT$ covariance matrix for all of the disturbances is $\boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \mathbf{I}$ and

$$\boldsymbol{\Omega}^{-1} = \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}.^{5} \tag{10-9}$$

The efficient estimator is generalized least squares.[6] The GLS estimator is

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

Denote the mnth element of $\boldsymbol{\Sigma}^{-1}$ by σ^{mn}. Expanding the **Kronecker products** produces

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \sigma^{11}\mathbf{X}_1'\mathbf{X}_1 & \sigma^{12}\mathbf{X}_1'\mathbf{X}_2 & \cdots & \sigma^{1M}\mathbf{X}_1'\mathbf{X}_M \\ \sigma^{21}\mathbf{X}_2'\mathbf{X}_1 & \sigma^{22}\mathbf{X}_2'\mathbf{X}_2 & \cdots & \sigma^{2M}\mathbf{X}_2'\mathbf{X}_M \\ & & \vdots & \\ \sigma^{M1}\mathbf{X}_M'\mathbf{X}_1 & \sigma^{M2}\mathbf{X}_M'\mathbf{X}_2 & \cdots & \sigma^{MM}\mathbf{X}_M'\mathbf{X}_M \end{bmatrix}^{-1} \begin{bmatrix} \sum_{m=1}^{M}\sigma^{1m}\mathbf{X}_1'\mathbf{y}_m \\ \sum_{m=1}^{M}\sigma^{2m}\mathbf{X}_2'\mathbf{y}_m \\ \vdots \\ \sum_{m=1}^{M}\sigma^{Mm}\mathbf{X}_M'\mathbf{y}_m \end{bmatrix}. \tag{10-10}$$

[5]See Appendix Section A.5.5.

[6]See Zellner (1962).

The asymptotic covariance matrix for the GLS estimator is the bracketed inverse matrix in (10-10).[7] All the results of Chapter 9 for the generalized regression model extend to this model.

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 noted two important special cases:

1. If the equations are *actually* unrelated—that is, if $\sigma_{mn} = 0$ for $m \neq n$—then there is obviously no payoff to GLS estimation of the full set of equations. Indeed, full GLS is equation by equation OLS.[8]
2. If the equations have **identical explanatory variables**—that is, if $\mathbf{X}_m = \mathbf{X}_n = \mathbf{X}$—then generalized least squares (GLS) is identical to equation by equation ordinary least squares (OLS). This case is common, notably in the capital asset pricing model in empirical finance (see the chapter Introduction) and in VAR models. A proof is considered in the exercises. This general result is lost if there are any restrictions on $\boldsymbol{\beta}$, either within or across equations. (The application in Example 10.2 is one of these cases.) The $\mathbf{X}$ matrices are identical, but there are cross-equation restrictions on the parameters, for example, in (10-4), $\beta_1 = \alpha_3$ and $\gamma_1 = \alpha_4$. Also, the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ for this case is given by the large inverse matrix in brackets in (10-10), which would be estimated by Est.Asy.Cov $[\hat{\boldsymbol{\beta}}_m, \hat{\boldsymbol{\beta}}_n] = \hat{\sigma}_{mn}(\mathbf{X}'\mathbf{X})^{-1}, m, n = 1, \ldots, M$, where $\hat{\sigma}_{mn} = \mathbf{e}_m'\mathbf{e}_n/T$. For the full set of estimators, Est.Asy.Cov$[\hat{\boldsymbol{\beta}}] = \hat{\boldsymbol{\Sigma}} \otimes (\mathbf{X}'\mathbf{X})^{-1}$.

In the more general case, with unrestricted correlation of the disturbances and different regressors in the equations, the extent to which GLS provides an improvement over OLS is complicated and depends on the data. Two propositions that apply generally are as follows:

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

10.2.3 FEASIBLE GENERALIZED LEAST SQUARES

The computation in (10-10) assumes that $\boldsymbol{\Sigma}$ is known, which, as usual, is unlikely to be the case. FGLS estimators based on the OLS residuals may be used.[10] A first step to estimate the elements of $\boldsymbol{\Sigma}$ uses

$$\hat{\sigma}_{mn} = s_{mn} = \mathbf{e}_m'\mathbf{e}_n/T. \quad \textbf{(10-11)}$$

[7]A robust covariance matrix along the lines of (10-8) could be constructed. However, note that the structure of $\boldsymbol{\Sigma} = \mathrm{E}[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t']$ has been used explicitly to construct the GLS estimator. The greater generality would be accommodated by assuming that $\mathrm{E}[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathbf{X}_t] = \boldsymbol{\Sigma}_t$ is not restricted, again, a form of heteroscedasticity robust covariance matrix. This extension is not standard in applications, however. [See Wooldridge (2010, pp. 173–176) for further development.]

[8]See also Kruskal (1968), Baltagi (1989), and Bartels and Fiebig (1992) for other cases where OLS equals GLS.

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

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

With

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

in hand, FGLS can proceed as usual.

The FGLS estimator requires inversion of the matrix **S** where the *mn*th element is given by (10-11). This matrix is $M \times M$. It is computed from the least squares residuals using

$$\mathbf{S} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{e}_t' = \frac{1}{T}\mathbf{E}'\mathbf{E}, \tag{10-13}$$

where $\mathbf{e}_t'$ is a $1 \times M$ vector containing all M residuals for the M equations at time t, placed as the tth row of the $T \times M$ matrix of residuals, **E**. The rank of this matrix cannot be larger than T. Note what happens if $M > T$. In this case, the $M \times M$ matrix has rank T, which is less than M, so it must be singular, and the FGLS estimator cannot be computed. In Example 10.1, we aggregate the 48 states into $M = 9$ regions. It would not be possible to fit a full model for the $M = 48$ states with only $T = 17$ observations. The data set is too short to obtain a positive definite estimate of $\boldsymbol{\Sigma}$.

10.2.4 TESTING HYPOTHESES

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

$$F[J, MT - K] = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})/J}{\hat{\boldsymbol{\varepsilon}}'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\varepsilon}}/(MT - K)}. \tag{10-14}$$

The computation uses the the unknown $\boldsymbol{\Omega}$. If we insert the estimator $\hat{\boldsymbol{\Omega}}$ based on (10-11) and use the result that the denominator in (10-14) converges to one in T (M is fixed) then, in large samples, the statistic will behave the same as

$$\hat{F} = \frac{1}{J}\left(\mathbf{R}\hat{\hat{\boldsymbol{\beta}}} - \mathbf{q}\right)\left\{\mathbf{R}\ \text{Est.Asy.Var.}\left[\hat{\hat{\boldsymbol{\beta}}}\right]\mathbf{R}'\right\}^{-1}\left(\mathbf{R}\hat{\hat{\boldsymbol{\beta}}} - \mathbf{q}\right). \tag{10-15}$$

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 $n \rightarrow \infty$. Therefore, an alternative test statistic that has a limiting chi-squared distribution with J degrees of freedom when the null hypothesis is true is

$$J\hat{F} = \left(\mathbf{R}\hat{\hat{\boldsymbol{\beta}}} - \mathbf{q}\right)'\left\{\mathbf{R}\ \text{Est.Asy.Var.}\left[\hat{\hat{\boldsymbol{\beta}}}\right]\mathbf{R}'\right\}^{-1}\left(\mathbf{R}\hat{\hat{\boldsymbol{\beta}}} - \mathbf{q}\right). \tag{10-16}$$

This is a Wald statistic that measures the distance between $\mathbf{R}\hat{\hat{\boldsymbol{\beta}}}$ and $\mathbf{q}$.

One hypothesis of particular interest is the **homogeneity or pooling restriction** of equal coefficient vectors in (10-2). The pooling restriction is that $\boldsymbol{\beta}_m = \boldsymbol{\beta}_M$, $i = 1, \ldots, M - 1$. Consistent with (10-15) and (10-16), we would form the hypothesis as

$$\mathbf{R}\boldsymbol{\beta} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & -\mathbf{I} \\ \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} & -\mathbf{I} \\ & & \cdots & & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & -\mathbf{I} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \cdots \\ \boldsymbol{\beta}_M \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_1 - \boldsymbol{\beta}_M \\ \boldsymbol{\beta}_2 - \boldsymbol{\beta}_M \\ \cdots \\ \boldsymbol{\beta}_{M-1} - \boldsymbol{\beta}_M \end{pmatrix} = \mathbf{0}. \tag{10-17}$$

This specifies a total of $(M - 1)K$ restrictions on the $MK \times 1$ parameter vector. Denote the estimated asymptotic covariance for $\left(\hat{\hat{\boldsymbol{\beta}}}_m, \hat{\hat{\boldsymbol{\beta}}}_n\right)$ as $\hat{\mathbf{V}}_{mn}$. The matrix in braces in (10-16) would have the typical $K \times K$ block,

$$\left\{\mathbf{R}\ \text{Est.Asy.Var.}\left[\hat{\hat{\boldsymbol{\beta}}}\right]\mathbf{R}'\right\}_{mn} = \hat{\mathbf{V}}_{mn} - \hat{\mathbf{V}}_{mM} - \hat{\mathbf{V}}_{Mn} + \hat{\mathbf{V}}_{MM}.$$

It is also of interest to assess statistically whether the off-diagonal elements of $\boldsymbol{\Sigma}$ are zero. If so, then the efficient estimator for the full parameter vector, absent within group 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 (10-1) and (10-2). With normally distributed disturbances, the standard trio of tests, Wald, **likelihood ratio**, and Lagrange multiplier, can be used. The Wald test is likely to be quite cumbersome. The likelihood ratio statistic for testing the null hypothesis that the matrix $\boldsymbol{\Sigma}$ in (10-3) is a diagonal matrix against the alternative that Σ is simply an unrestricted positive definite matrix would be

$$\lambda_{LR} = T[\ln|\mathbf{S}_0| - \ln|\mathbf{S}_1|], \tag{10-18}$$

where $\mathbf{S}_1$ is the residual covariance matrix defined in (10-12) (without a degrees of freedom correction). The residuals are computed using maximum likelihood estimates of the parameters, not FGLS.[11] Under the null hypothesis, the model would be efficiently estimated by individual equation OLS, so

$$\ln|\mathbf{S}_0| = \sum_{m=1}^{M} \ln(\mathbf{e}_m'\mathbf{e}_m/T).$$

The statistic would be used for a chi-squared test with $M(M - 1)/2$ degrees of freedom. The Lagrange multiplier statistic developed by Breusch and Pagan (1980) is

$$\lambda_{LM} = T\sum_{m=2}^{M}\sum_{n=1}^{m-1} r_{mn}^2, \tag{10-19}$$

based on 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. Alternative approaches that have been suggested, such as the LR test in (10-18), are based on the "excess variation," $(\hat{\boldsymbol{\Sigma}}_0 - \hat{\boldsymbol{\Sigma}}_1)$.[12]

[11] 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 (10-10) and (10-12), until the estimates are no longer changing.

[12] See, for example, Johnson and Wichern (2005, p. 424).

10.2.5 THE POOLED MODEL

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

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

results. Collecting the T observations for group m, we obtain

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

For all M groups,

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_M \end{bmatrix}\boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{10-20}$$

where

$$\begin{aligned} E[\boldsymbol{\varepsilon}_i \,|\, \mathbf{X}] &= \mathbf{0}, \\ E[\boldsymbol{\varepsilon}_m\boldsymbol{\varepsilon}'_n \,|\, \mathbf{X}] &= \sigma_{mn}\mathbf{I}, \end{aligned} \tag{10-21}$$

or

$$E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \boldsymbol{\Sigma} \otimes \mathbf{I}.$$

The generalized least squares estimator under this assumption is

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= [\mathbf{X}'(\boldsymbol{\Sigma} \otimes \mathbf{I})^{-1}\mathbf{X}]^{-1}[\mathbf{X}'(\boldsymbol{\Sigma} \otimes \mathbf{I})^{-1}\mathbf{y}] \\ &= \left[\sum_{m=1}^{M}\sum_{n=1}^{M}\sigma^{mn}\mathbf{X}'_m\mathbf{X}_n\right]^{-1}\left[\sum_{m=1}^{M}\sum_{n=1}^{M}\sigma^{mn}\mathbf{X}'_m\mathbf{y}_n\right]. \end{aligned} \tag{10-22}$$

The FGLS estimator can be computed using (10-11), where $\mathbf{e}_m$ would be a subvector of the pooled OLS residual vector using all MT observations.

Example 10.1 A Regional Production Model for Public Capital

Munnell (1990) proposed a model 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,

$$\begin{aligned} \ln gsp_{mt} = \alpha_m &+ \beta_{1m}\ln pc_{mt} + \beta_{2m}\ln hwy_{mt} + \beta_{3m}\ln water_{mt} + \beta_{4m}\ln util_{mt} \\ &+ \beta_{5m}\ln emp_{mt} + \beta_{6m}\, unemp_{mt} + \varepsilon_{mt}, \end{aligned}$$

where

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

The data, measured for the 48 contiguous states in the U.S. (excluding Alaska and Hawaii) and years 1970–1986 are given in Appendix Table F10.1. We will aggregate the data for the 48 states into nine regions consisting of the following groups of states (the state codes appear in the data file):

Gulf States = GF = AL, FL, LA, MS,
Southwest = SW = AZ, NV, NM, TX, UT,
West Central = WC = CA, OR, WA,
Mountain = MT = CO, ID, MT, ND, SD, WY,
Northeast = NE = CT, ME, MA, NH, RI, VT,
Mid Atlantic = MA = DE, MD, NJ, NY, PA, VA,
South = SO = GA, NC, SC, TN, WV, AR,
Midwest = MW = IL, IN, KY, MI, MN, OH, WI,
Central = CN = IA, KS, MO, NE, OK.

This defines a nine-equation model. Note that with only 17 observations per state, it is not possible to fit the unrestricted 48-equation model. This would be a case of the short rank problem noted at the end of Section 10.2.2. The calculations for the data setup are are described in Application 1 at the end of this chapter, where the reader is invited to replicate the computations and fill in the omitted parts of Table 10.3.

We initially estimated the nine equations of the regional productivity model separately by OLS. The OLS estimates are shown in Table 10.1. (For brevity, the estimated standard errors are not shown.)

The correlation matrix for the OLS residuals is shown in Table 10.2.

TABLE 10.1 Estimates of Seemingly Unrelated Regression Equations

Region		α	β_1	β_2	β_3	β_4	β_5	β_6	R^2
GF	OLS	11.570	0.002	−2.028	0.101	1.358	0.805	−0.007	0.997
	FGLS	12.310	−0.201	−1.886	0.178	1.190	0.953	−0.003	
SW	OLS	3.028	0.164	−0.075	−0.169	0.637	0.362	−0.017	0.998
	FGLS	4.083	0.077	−0.131	−0.136	0.522	0.539	−0.156	
WC	OLS	3.590	0.295	0.174	−0.226	−0.215	0.917	−0.008	0.994
	FGLS	1.960	0.170	0.132	−0.347	0.895	1.070	−0.006	
MT	OLS	6.378	−0.153	−0.123	0.306	−0.533	1.344	0.005	0.999
	FGLS	3.463	−0.115	0.180	0.262	−0.330	1.079	−0.002	
NE	OLS	−13.730	−0.020	0.661	−0.969	−0.107	3.380	0.034	0.985
	FGLS	−12.294	−0.118	0.934	−0.557	−0.290	2.494	0.020	
MA	OLS	−22.855	−0.378	3.348	−0.264	−1.778	2.637	0.026	0.986
	FGLS	−18.616	−0.311	3.060	−0.109	−1.659	2.186	0.018	
SO	OLS	3.922	0.043	−0.773	−0.035	0.137	1.665	0.008	0.994
	FGLS	3.162	−0.063	−0.641	−0.081	0.281	1.620	0.008	
MW	OLS	−9.111	0.233	1.604	0.717	−0.356	−0.259	−0.034	0.989
	FGLS	−9.258	0.096	1.612	0.694	−0.340	−0.062	−0.031	
CN	OLS	−5.621	0.386	1.267	0.546	−0.108	−0.475	−0.313	0.995
	FGLS	−3.405	0.295	0.934	0.539	0.003	−0.321	−0.030	

TABLE 10.2 Correlations of OLS Residuals

	GF	*SW*	*WC*	*MT*	*NE*	*MA*	*SO*	*MW*	*CN*
GF	1								
SW	0.173	1							
WC	0.447	0.697	1						
MT	−0.547	−0.290	−0.537	1					
NE	0.525	0.489	0.343	−0.241	1				
MA	0.425	0.132	0.130	−0.322	0.259	1			
SO	0.763	0.314	0.505	−0.351	0.783	0.388	1		
MW	0.167	0.565	0.574	−0.058	0.269	−0.037	0.366	1	
CN	0.325	0.119	0.037	0.091	0.200	0.713	0.350	0.298	1

The correlations are large enough to suggest that there is substantial correlation of the disturbances across regions. The LM statistic in (10-19) for testing the hypothesis that the covariance matrix of the disturbances is diagonal equals 103.1 with $8(9)/2 = 36$ degrees of freedom. The critical value from the chi-squared table is 50.998, so the null hypothesis that $\sigma_{mn} = 0$ (or $\rho_{mn} = 0$) for all $m \neq n$, that is, that the seemingly unrelated regressions are actually unrelated, is rejected on this basis. Table 10.1 also presents the FGLS estimates of the model parameters. These are computed in two steps, with the first-step OLS results producing the estimate of $\boldsymbol{\Sigma}$ for FGLS. The correlations in Table 10.2 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 and heteroscedasticity. A comparison of some of the estimates for the main capital and labor coefficients appears in Table 10.3. The estimates themselves are comparable. But the estimated standard errors for the FGLS coefficients are roughly half as large as the corresponding OLS values. This suggests a large gain in efficiency from using GLS rather than OLS.

The pooling restriction is formulated as

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

TABLE 10.3 Comparison of OLS and FGLS Estimates*

	$\boldsymbol{\beta}_1$		$\boldsymbol{\beta}_5$	
Region	***OLS***	***FGLS***	***OLS***	***FGLS***
GF	0.002 (0.301)	−0.201 (0.142)	0.805 (0.159)	0.953 (0.085)
SW	0.164 (0.166)	0.077 (0.086)	0.362 (0.165)	0.539 (0.085)
WC	0.295 (0.205)	0.170 (0.092)	0.917 (0.377)	1.070 (0.171)
MT	−0.153 (0.084)	−0.115 (0.048)	1.344 (0.188)	1.079 (0.105)
NE	−0.020 (0.286)	−0.118 (0.131)	3.380 (1.164)	2.494 (0.479)
MA	−0.378 (0.167)	−0.311 (0.081)	2.673 (1.032)	2.186 (0.448)
SO	0.043 (0.279)	−0.063 (0.104)	1.665 (0.414)	1.620 (0.185)
MW	0.233 (0.206)	0.096 (0.102)	−0.259 (0.303)	−0.062 (0.173)
CN	0.386 (0.211)	0.295 (0.090)	−0.475 (0.259)	−0.321 (0.169)
Pooled	0.260 (0.017)	0.254 (0.006)	0.330 (0.030)	0.343 (0.001)

*Estimates of Capital (β_1) and Labor (β_5) coefficients. Estimated standard errors in parentheses.

The **R** matrix for this hypothesis is shown in (10-17). The test statistic is in (10-16). For our model with nine equations and seven parameters in each, the null hypothesis imposes (9-1)7 = 56 restrictions. The computed test statistic is 6092.5, which is far larger than the critical value from the chi-squared table, 74.468. So, the hypothesis of homogeneity is rejected. Part of the pooled estimator is shown in Table 10.3. The benefit of the restrictions on the estimator can be seen in the much smaller standard errors in every case compared to the separate estimators. If the hypothesis that all the coefficient vectors were the same were true, the payoff to using that information would be obvious. Because the hypothesis is rejected, that benefit is less clear, as now the pooled estimator does not consistently estimate any of the individual coefficient vectors.

10.3 SYSTEMS OF DEMAND EQUATIONS: SINGULAR SYSTEMS

Many of the applications of the seemingly unrelated regression model have estimated **systems of demand equations**, either commodity demands, factor demands, or factor share equations in studies of production. Each is merely a particular application of the model of Section 10.2. But some special problems arise in these settings. First, the parameters of the systems are usually constrained across the equations. This usually takes the form of parameter equality constraints across the equations, such as the symmetry assumption in production and cost models—see (10-32) and (10-33).[13] A second feature of many of these models is that the disturbance covariance matrix $\boldsymbol{\Sigma}$ is singular, which would seem to preclude GLS (or FGLS).

10.3.1 COBB–DOUGLAS COST FUNCTION

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

$$Q = \alpha_0 \prod_{m=1}^{M} x_m^{\alpha_m}.$$

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

$$\Lambda = \alpha_0 \prod_{m=1}^{M} x_m^{\alpha_m} + \lambda(C - \mathbf{p}'\mathbf{x}),$$

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

$$\frac{\partial \Lambda}{\partial x_m} = \frac{\alpha_m Q}{x_m} - \lambda p_m = 0 \quad \text{and} \quad \frac{\partial \Lambda}{\partial \lambda} = C - \mathbf{p}'\mathbf{x} = 0.$$

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

$$\sum_{m=1}^{M} p_m x_m = \sum_{m=1}^{M} \frac{\alpha_m Q}{\lambda}.$$

The cost share allocated to the mth factor is

[13]See Silver and Ali (1989) for a discussion of testing symmetry restrictions.

$$\frac{p_m x_m}{\sum_{m=1}^{M} p_m x_m} = \frac{\alpha_m}{\sum_{m=1}^{M} \alpha_m} = \beta_m. \tag{10-23}$$

The full model is[14]

$$\begin{aligned} \ln C &= \beta_0 + \beta_q \ln Q + \sum_{m=1}^{M} \beta_m \ln p_m + \varepsilon_c, \\ s_m &= \beta_m + \varepsilon_m, m = 1, \ldots, M. \end{aligned} \tag{10-24}$$

Algebraically, $\sum_{m=1}^{M} \beta_m = 1$ and $\sum_{m=1}^{M} s_m = 1$. (This is the cost function analysis begun in Example 6.17. We will return to that application below.) The cost shares will also sum identically to one in the data. It therefore follows that $\sum_{m=1}^{M} \varepsilon_m = 0$ 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 $\sum_{m=1}^{M} \beta_m = 1$ states that the cost function must be homogeneous of degree one in the prices. If we impose the constraint

$$\beta_M = 1 - \beta_1 - \beta_2 - \cdots - \beta_{M-1}, \tag{10-25}$$

then the system is reduced to a nonsingular one,

$$\begin{aligned} \ln\left(\frac{C}{p_M}\right) &= \beta_0 + \beta_q \ln Q + \sum_{m=1}^{M-1} \beta_m \ln\left(\frac{p_m}{p_M}\right) + \varepsilon_c, \\ s_m &= \beta_m + \varepsilon_m, \quad m = 1, \ldots, M - 1. \end{aligned}$$

This system provides estimates of β_0, β_q, and $\beta_1, \ldots, \beta_{M-1}$. The last parameter is estimated using (10-25). It is immaterial which factor is chosen as the numeraire; FGLS will be **invariant** to which factor is chosen.

Example 10.2 Cobb–Douglas Cost Function

Nerlove's (1963) study of the electric power industry that we examined in Example 6.6 provides an application of the Cobb–Douglas cost function model. His ordinary least squares estimates of the parameters were listed in Example 6.6. 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 F6.2 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 10.4.

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

Least squares estimates of the Cobb–Douglas cost function are given in the first column. The coefficient on capital is negative. Because $\beta_m = \beta_q \partial \ln Q/\partial \ln x_m$—that is, a positive multiple of the output elasticity of the mth factor—this finding is troubling. 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 (10-20),

$$\mathbf{y} = \begin{bmatrix} \ln(C/P_f) \\ s_k \\ s_l \end{bmatrix} = \begin{bmatrix} \mathbf{i} & \ln Q & \ln(P_k/P_f) & \ln(P_l/P_f) \\ \mathbf{0} & \mathbf{0} & \mathbf{i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{i} \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_q \\ \beta_k \\ \beta_l \end{pmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_c \\ \boldsymbol{\varepsilon}_k \\ \boldsymbol{\varepsilon}_l \end{bmatrix}.$$

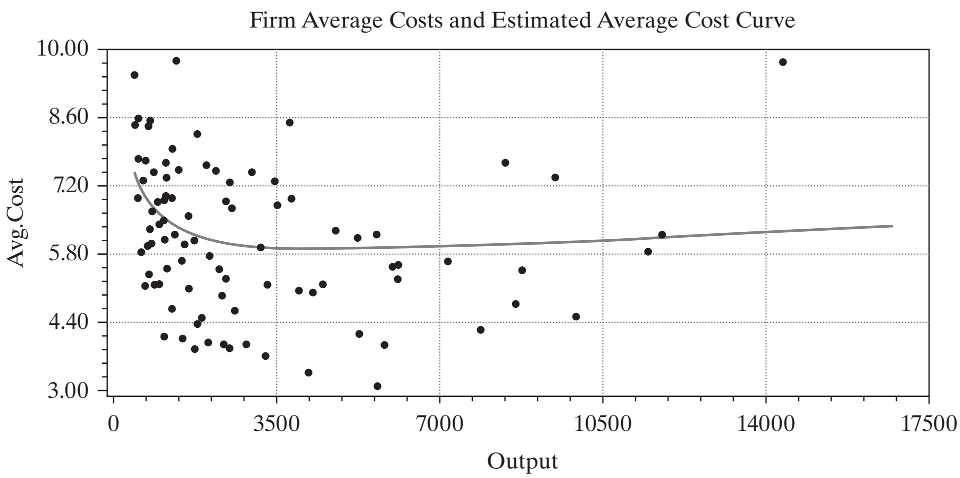
Note this formulation imposes the restrictions $\beta_1 = \alpha_3$ and $\gamma_1 = \alpha_4$ on (10-4). There are 3(145) = 435 observations in the data matrices. The estimator is then FGLS, as shown in (10-22). An additional column is added for the log quadratic model. Two things to note are the dramatically 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 FGLS estimates are given in the second and fourth sets of results.

The quadratic output term gives the average 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 $Q* = \exp[(1 - \beta_q)/(2\beta_{qq})]$. Using the FGLS estimates, this value is $Q* = 4{,}669$. (Application 5 considers using the delta method to construct a confidence interval for Q^*.) About 85% 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.

TABLE 10.4 Cost Function Estimates (Estimated standard errors in parentheses)

		Ordinary Least Squares		*Constrained Feasible GLS*	
Constant	β_0	−4.686	−3.764	−7.069	−5.707
		(0.885)	(0.702)	(0.107)	(0.165)
ln *Output*	β_q	0.721	0.153	0.766	0.239
		(0.0174)	(0.0618)	(0.0154)	(0.0587)
$\ln^2$ *Output*	β_{qq}		0.0505		0.0451
			(0.0054)		(0.00508)
$\ln P_{capital}$	β_k	−0.0085	0.0739	0.424	0.425
		(0.191)	(0.150)	(0.00946)	(0.00943)
$\ln P_{labor}$	β_l	0.594	0.481	0.106	0.106
		(0.205)	(0.161)	(0.00386)	(0.00380)
$\ln P_{fuel}$	β_f	0.414	0.445	0.470	0.470
		(0.0989)	(0.0777)	(0.0101)	(0.0100)

FIGURE 10.1 Predicted Average Costs.



10.3.2 FLEXIBLE FUNCTIONAL FORMS: THE TRANSLOG COST FUNCTION

The classic paper by Arrow et al. (1961) called into question the inherent restriction of the popular Cobb–Douglas model that all elasticities of factor substitution are equal to one. Researchers have since developed numerous **flexible functions** that allow substitution to be unrestricted.[15] Similar strands of literature have appeared in the analysis of commodity demands.[16] In this section, we examine in detail a specific 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 $x_m* = x_m(Q, \mathbf{p})$. The total cost of production is given by the cost function,

$$C = \sum_{m=1}^{M} p_m x_m(Q, \mathbf{p}) = C(Q, \mathbf{p}). \qquad \textbf{(10-26)}$$

If there are **constant returns to scale**, then it can be shown that $C = Qc(\mathbf{p})$ or $C/Q = c(\mathbf{p})$, where $c(\mathbf{p})$ is the per unit or average cost function.[17] The cost-minimizing factor demands are obtained by applying **Shephard's lemma (1970)**, which states that if $C(Q, \mathbf{p})$ gives the minimum total cost of production, then the cost-minimizing set of factor demands is given by

$$x_m^* = \frac{\partial C(Q, \mathbf{p})}{\partial p_m}. \qquad \textbf{(10-27)}$$

---

[15]See, in particular, Berndt and Christensen (1973).

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

[17]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.

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

$$s_m^* = \frac{\partial \ln C(Q, \mathbf{p})}{\partial \ln p_m} = \frac{p_m}{C}\frac{\partial C(Q, \mathbf{p})}{\partial p_m} = \frac{p_m x_m^*}{C}. \quad \textbf{(10-28)}$$

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

$$s_m^* = \frac{\partial \ln c(\mathbf{p})}{\partial \ln p_m}. \quad \textbf{(10-29)}$$

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

$$\theta_{mn} = \frac{c(\partial^2 c/\partial p_m \partial p_n)}{(\partial c/\partial p_m)(\partial c/\partial p_n)}$$

and

$$\eta_m = s_m \theta_{mm}.$$

By suitably parameterizing the cost function (10-26) and the cost shares (10-29), we obtain an M or $M + 1$ equation econometric model that can be used to estimate these quantities.

The transcendental logarithmic or **translog function** is the most frequently used flexible function in empirical work.[18] By expanding $\ln c(\mathbf{p})$ in a second-order Taylor series about the point $\ln(\mathbf{p}) = \mathbf{0}$, we obtain

$$\ln c \approx \beta_0 + \sum_{m=1}^{M}\left(\frac{\partial \ln c}{\partial \ln p_m}\right)\log p_m + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M}\left(\frac{\partial^2 \ln c}{\partial \ln p_m \,\partial \ln p_n}\right)\ln p_m \ln p_n, \quad \textbf{(10-30)}$$

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

$$\begin{aligned}\ln c = \beta_0 &+ \beta_1 \ln p_1 + \cdots + \beta_M \ln p_M + \delta_{11}\left(\frac{1}{2}\ln^2 p_1\right) + \delta_{12}\ln p_1 \ln p_2 \\ &+ \delta_{22}\left(\frac{1}{2}\ln^2 p_2\right) + \cdots + \delta_{MM}\left(\frac{1}{2}\ln^2 p_M\right).\end{aligned} \quad \textbf{(10-31)}$$

This is the translog cost function. If δ_{mn} equals zero, then it reduces to the Cobb–Douglas function in Section 10.3.1. The cost shares are given by

$$\begin{aligned}s_1 &= \frac{\partial \ln c}{\partial \ln p_1} = \beta_1 + \delta_{11}\ln p_1 + \delta_{12}\ln p_2 + \cdots + \delta_{1M}\ln p_M, \\ s_2 &= \frac{\partial \ln c}{\partial \ln p_2} = \beta_2 + \delta_{21}\ln p_1 + \delta_{22}\ln p_2 + \cdots + \delta_{2M}\ln p_M, \\ &\vdots \\ s_M &= \frac{\partial \ln c}{\partial \ln p_M} = \beta_M + \delta_{M1}\ln p_1 + \delta_{M2}\ln p_2 + \cdots + \delta_{MM}\ln p_M.\end{aligned} \quad \textbf{(10-32)}$$

[18] The function was proposed in a series of papers by Berndt, Christensen, Jorgenson, and Lau, including Berndt and Christensen (1973) and Christensen et al. (1975).

The theory implies a number of restrictions on the parameters. The matrix of second derivatives must be symmetric (by Young's theorem for continuous functions). The cost function must be linearly homogeneous in the factor prices. This implies $\Sigma_{m=1}^{M}(\partial \ln c(\mathbf{p})/\partial \ln p_m) = 1$. This implies the adding-up restriction, $\Sigma_{m=1}^{M} s_m = 1$. Together, these imply the following set of cross-equation restrictions:

$$\begin{aligned} \delta_{mn} &= \delta_{nm} && \text{(symmetry)}, \\ \sum_{m=1}^{M} \beta_m &= 1 && \text{(linear homogeneity)}, \\ \sum_{m=1}^{M} \delta_{mn} &= \sum_{n=1}^{M} \delta_{mn} = 0. \end{aligned} \tag{10-33}$$

The system of **share equations** in (10-32) produces a seemingly unrelated regressions model that can be used to estimate the parameters of the model.[19] To make the model operational, we must impose the restrictions in (10-33) and solve the problem of **singularity of the disturbance covariance matrix** of the share equations. The first is accomplished by dividing the first $M - 1$ prices by the Mth, thus eliminating the last term in each row and column of the parameter matrix. As in the Cobb–Douglas model, we obtain a nonsingular system by dropping the Mth share equation. For the translog cost function, the elasticities of substitution are particularly simple to compute once the parameters have been estimated,

$$\theta_{mn} = \frac{\delta_{mn} + s_m s_n}{s_m s_n}, \quad \theta_{mm} = \frac{\delta_{mm} + s_m(s_m - 1)}{s_m^2}. \tag{10-34}$$

These elasticities will differ at every data point. It is common to compute them at some central point such as the means of the data.[20] The factor-specific demand elasticities are then computed using $\eta_m = s_m \theta_{mm}$.

Example 10.3 A Cost Function for U.S. Manufacturing

A number of studies using the translog methodology have used a four-factor model, with capital K, labor L, energy E, and materials M, the factors of production. Among the studies to employ this methodology was Berndt and Wood's (1975) estimation of a translog cost function for the U.S. manufacturing sector. The three factor shares used to estimate the model are

$$\begin{aligned} s_K &= \beta_K + \delta_{KK} \ln\left(\frac{p_K}{p_M}\right) + \delta_{KL} \ln\left(\frac{p_L}{p_M}\right) + \delta_{KE} \ln\left(\frac{p_E}{p_M}\right), \\ s_L &= \beta_L + \delta_{KL} \ln\left(\frac{p_K}{p_M}\right) + \delta_{LL} \ln\left(\frac{p_L}{p_M}\right) + \delta_{LE} \ln\left(\frac{p_E}{p_M}\right), \\ s_E &= \beta_E + \delta_{KE} \ln\left(\frac{p_K}{p_M}\right) + \delta_{LE} \ln\left(\frac{p_L}{p_M}\right) + \delta_{EE} \ln\left(\frac{p_E}{p_M}\right). \end{aligned}$$

[19]The system of factor share equations estimates all of the parameters in the model except for the overall constant term, β_0. The cost function can be omitted from the model. Without the assumption of constant returns to scale, however, the cost function will contain parameters of interest that do not appear in the share equations. In this case, one would want to include it in the equation system. See Christensen and Greene (1976) for an application.

[20]They will also be highly nonlinear functions of the parameters and the data. A method of computing asymptotic standard errors for the estimated elasticities is presented in Anderson and Thursby (1986). Krinsky and Robb (1986, 1990, 1991). (See also Section 15.3.) proposed their method as an alternative approach to this computation.

Berndt and Wood's data are reproduced in Appendix Table F10.2. Constrained FGLS estimates of the parameters presented in Table 10.4 were obtained by constructing the pooled regression in (10-20) with data matrices

$$\mathbf{y} = \begin{bmatrix} \mathbf{s_K} \\ \mathbf{s_L} \\ \mathbf{s_E} \end{bmatrix}, \tag{10-35}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{i} & \mathbf{0} & \mathbf{0} & \ln \mathbf{P_K/P_M} & \ln \mathbf{P_L/P_M} & \ln \mathbf{P_E/P_M} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{i} & \mathbf{0} & \mathbf{0} & \ln \mathbf{P_K/P_M} & \mathbf{0} & \ln \mathbf{P_L/P_M} & \ln \mathbf{P_K/P_M} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{i} & \mathbf{0} & \mathbf{0} & \ln \mathbf{P_K/P_M} & \mathbf{0} & \ln \mathbf{P_L/P_M} & \ln \mathbf{P_E/P_M} \end{bmatrix},$$

$$\boldsymbol{\beta}' = (\beta_K, \beta_L, \beta_E, \delta_{KK}, \delta_{KL}, \delta_{KE}, \delta_{LL}, \delta_{LE}, \delta_{EE}).$$

Estimates are then obtained by iterating the two-step procedure in (10-11) and (10-22).[21] The parameters not estimated directly in (10-35) are computed using (10-33). The implied estimates of the elasticities of substitution and demand elasticities for 1959 (the central year in the data) are given in TABLE 10.5 using the fitted cost shares and the estimated parameters in (10-34). The departure from the Cobb–Douglas model with unit elasticities is substantial. For example, the results suggest almost no substitutability between energy and labor and some complementarity between capital and energy.

The underlying theory requires that the cost function satisfy three regularity conditions, homogeneity of degree one in the input prices, monotonicity in the prices, and quasiconcavity. The first of these is imposed by (10-33), which we built into the model. The second is obtained if all of the fitted cost shares are positive, which we have verified at every observation. The third requires that the matrix,

$$\mathbf{F}_t = \boldsymbol{\Delta} - diag(\mathbf{s}_t) + \mathbf{s}_t\mathbf{s}_t',$$

TABLE 10.5 Parameter Estimates for Aggregate Translog Cost Function (Standard errors in parentheses)

	Constant	*Capital*	*Labor*	*Energy*	*Materials*
Capital	0.05689	0.02949	−0.00005	−0.01067	−0.01877*
	(0.00135)	(0.00580)	(0.00385)	(0.00339)	(0.00971)
Labor	0.25344		0.07543	−0.00476	−0.07063*
	(0.00223)		(0.00676)	(0.00234)	(0.01060)
Energy	0.04441			0.01835	−0.00294*
	(0.00085)			(0.00499)	(0.00800)
Materials	0.64526*				0.09232*
	(0.00330)				(0.02247)

*Derived using (10-33).

[21]The estimates do not match those reported by Berndt and Wood. To purge their data of possible correlation with the disturbances, they first regressed the prices on 10 exogenous macroeconomic variables, such as U.S. population, government purchases of labor services, real exports of durable goods and U.S. tangible capital stock, and then based their analysis on the fitted values. The estimates given here are, in general, quite close to theirs. For example, their estimates of the constants in Table 10.5 are 0.60564, 0.2539, 0.0442, and 0.6455. Berndt and Wood's estimate of θ_{EL} for 1959 is 0.64 compared to ours in Table 10.5 of 0.60564.

TABLE 10.6 Estimated Elasticities

	Capital	*Labor*	*Energy*	*Materials*
		Cost Shares for 1959		
Fitted	0.05640	0.27452	0.04389	0.62519
Actual	0.06185	0.27303	0.04563	0.61948
		Implied Elasticities of Substitution, 1959		
Capital	−7.4612			
Labor	0.99691	−1.64179		
Energy	−3.31133	0.60533	−12.2566	
Materials	0.46779	0.58848	0.89334	−0.36331
		Implied Own Price Elasticities		
	−0.420799	−0.45070	−0.53793	−0.22714

be negative semidefinite, where $\boldsymbol{\Delta}$ is the symmetric matrix of coefficients on the quadratic terms in Table 10.5 and $\mathbf{s}_t$ is the vector of factor shares. This condition can be checked at each observation by verifying that the characteristic roots of $\mathbf{F}_t$ are all nonpositive. For the 1959 data, the four characteristic roots are (0, −0.00152, −0.06277, −0.23514). The results for the other years are similar. The estimated cost function satisfies the theoretical regularity conditions.

10.4 SIMULTANEOUS EQUATIONS MODELS

The seemingly unrelated regression model,

$$y_{mt} = \mathbf{x}_{mt}'\boldsymbol{\beta}_m + \varepsilon_{mt},$$

derives from a set of regression equations that are related through the disturbances. The regressors, $\mathbf{x}_{mt}$, are *exogenous* and can vary for reasons that are not explained within the model. Thus, the coefficients are directly interpretable as partial or causal effects and can be estimated by least squares or other methods that are based on the conditional mean functions, $E[y_{mt}|\mathbf{x}_{mt}] = \mathbf{x}_{mt}'\boldsymbol{\beta}$. In the market equilibrium model suggested in the Introduction,

$$\begin{aligned} Q_{Demand} &= \alpha_1 + \alpha_2 Price + \alpha_3 Income + \mathbf{d}'\boldsymbol{\alpha} + \varepsilon_{Demand}, \\ Q_{Supply} &= \beta_1 + \beta_2 Price + \beta_3 FactorPrice + \mathbf{s}'\boldsymbol{\beta} + \varepsilon_{Supply}, \\ Q_{Equilibrium} &= Q_{Demand} = Q_{Supply}, \end{aligned}$$

neither of the two market equations is a conditional mean. The partial equilibrium experiment of changing the equilibrium price and inducing a change in the equilibrium quantity in the hope of eliciting an estimate of the demand elasticity, α_2 (or supply elasticity, β_2), makes no sense. The model is of the joint determination of quantity *and* price. Price changes when the market equilibrium changes, but that is induced by changes in other factors, such as changes in incomes or other variables that affect the supply function. Nonetheless, the elasticities of demand and supply, α_2 and β_2, are of interest, and do have a causal interpretation in the context of the model. This section considers the theory and methods that apply for estimation and analysis of systems of interdependent equations.

As we saw in Example 8.4, least squares regression of observed equilibrium quantities on price and the other factors will compute an ambiguous mixture of the supply and demand functions. The result follows from the *endogeneity* of *Price* in either equation. Simultaneous equations models arise in settings such as this one, in which the set of equations are interdependent. Simultaneous equations models will fit in the framework developed in Chapter 8, where we considered equations in which some of the right-hand-side variables are endogenous—that is, correlated with the disturbances. The substantive difference at this point is the source of the endogeneity. In our treatments in Chapter 8, endogeneity arose, for example, in the models of omitted variables, measurement error, or endogenous treatment effects, essentially as an unintended deviation from the assumptions of the linear regression model. In the simultaneous equations framework, endogeneity is a fundamental part of the specification. This section will consider the issues of specification and estimation in systems of simultaneous equations. We begin in Section 10.4.1 with a development of a general framework for the analysis and a statement of some fundamental issues. Section 10.4.2 presents the simultaneous equations model as an extension of the seemingly unrelated regressions model in Section 10.2. The ultimate objective of the analysis will be to learn about the model coefficients. The issue of whether this is even possible is considered in Section 10.4.3, where we develop the issue of identification. Once the identification question is settled, methods of estimation and inference are presented in Sections 10.4.4 and 10.4.5.

Example 10.4. Reverse Causality and Endogeneity in Health

As we examined in Chapter 8, endogeneity arises from several possible sources. The case considered in this chapter is simultaneity, sometimes labeled *reverse causality.* Consider a familiar modeling framework in health economics, the “health production function” (see Grossman (1972)), in which we might model health outcomes as

$$\textit{Health} = f(\textit{Income}, \textit{Education}, \textit{Health Care}, \textit{Age}, \ \ldots, \varepsilon_H = \text{other factors}).$$

It is at least debatable whether this can be treated as a regression. For any individual, arguably, lower incomes are associated with lower results for health. But which way does the “causation run?” It may also be that variation in health is a driver of variation in income. A natural companion might appear

$$\textit{Income} = g(\textit{Health}, \textit{Education}, \ \ldots, \varepsilon_I = \text{labor market factors}).$$

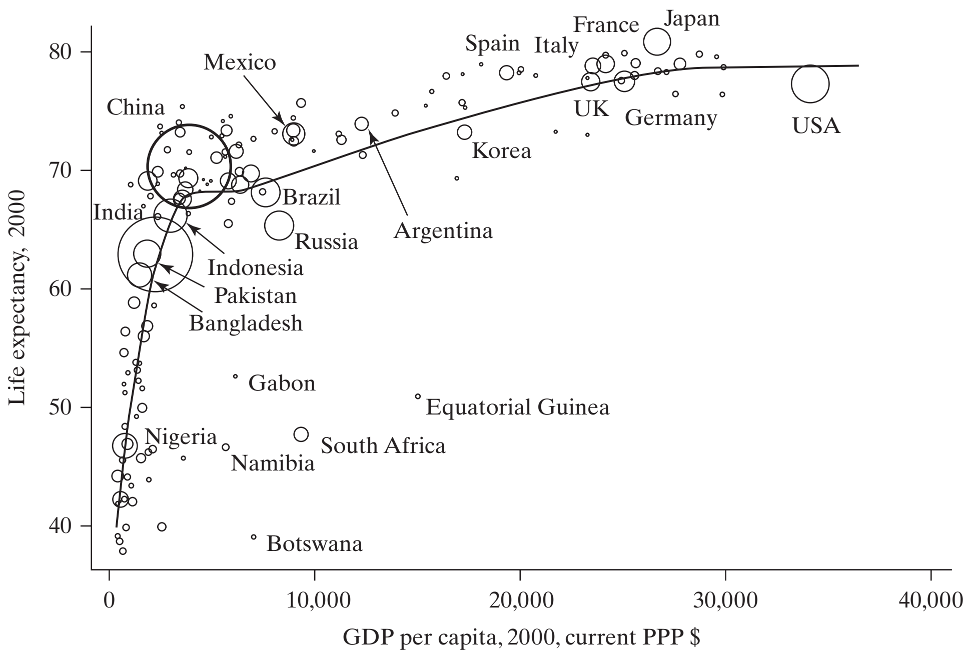
The causal effect of income on health could, in principle, be examined through the experiment of varying income, assuming that external factors such as labor market conditions could be driving the change in income. But, in the second equation, we could likewise be interested in how variation in health outcomes affect incomes. The idea is similarly complicated at the aggregate level. Deaton’s (2003) updated version of the “Preston Curve” (1978) in Figure 10.2 suggests covariation between health (life expectancy) and income (per capita GDP) for a group of countries. Which variable is driving which is part of a longstanding discussion.

10.4.1 SYSTEMS OF EQUATIONS

Consider a simplified version of the equilibrium model,

$$\begin{aligned} \text{demand equation:} \quad & q_{d,t} = \alpha_1 p_t + \alpha_2 x_t + \varepsilon_{d,t}, \\ \text{supply equation:} \quad & q_{s,t} = \beta_1 p_t + \varepsilon_{s,t}, \\ \text{equilibrium condition:} \quad & q_{d,t} = q_{s,t} = q_t. \end{aligned}$$

FIGURE 10.2 Updated Preston Curve.



These equations are **structural equations** in that they are derived from theory and each purports to describe a particular aspect of the economy. Because the model is one of the joint determination of price and quantity, they are labeled **jointly dependent** or **endogenous** variables. Income, x, is assumed to be determined outside of the model, which makes it **exogenous**. The disturbances are added to the usual textbook description to obtain an **econometric model**. All three equations are needed to determine the equilibrium price and quantity, so the system is **interdependent**. Finally, because an equilibrium solution for price and quantity in terms of income and the disturbances is, indeed, implied (unless α_1 equals β_1), the system is said to be a **complete system of equations**. As a general rule, it is not possible to estimate all the parameters of incomplete systems. (It may be possible to estimate some of them, as will turn out to be the case with this example).

Suppose that interest centers on estimating the demand elasticity α_1. For simplicity, assume that ε_d and ε_s are well behaved, classical disturbances with

$$\begin{aligned} E[\varepsilon_{d,t}|x_t] &= E[\varepsilon_{s,t}|x_t] = 0, \\ E[\varepsilon_{d,t}^2|x_t] &= \sigma_d^2, \\ E[\varepsilon_{s,t}^2|x_t] &= \sigma_s^2, \\ E[\varepsilon_{d,t}\varepsilon_{s,t}|x_t] &= 0. \end{aligned}$$

All variables are mutually uncorrelated with observations at different time periods. Price, quantity, and income are measured in logarithms in deviations from their sample means. Solving the equations for p and q in terms of x, ε_d, and ε_s produces the **reduced form** of the model,

$$p = \frac{\alpha_2 x}{\beta_1 - \alpha_1} + \frac{\varepsilon_d - \varepsilon_s}{\beta_1 - \alpha_1} = \pi_1 x + v_1,$$
$$q = \frac{\beta_1 \alpha_2 x}{\beta_1 - \alpha_1} + \frac{\beta_1 \varepsilon_d - \alpha_1 \varepsilon_s}{\beta_1 - \alpha_1} = \pi_2 x + v_2. \quad \textbf{(10-36)}$$

(Note the role of the "completeness" requirement that α_1 not equal β_1. This means that the two lines are not parallel.) It follows that $\text{Cov}[p, \varepsilon_d] = \sigma_d^2/(\beta_1 - \alpha_1)$ and $\text{Cov}[p, \varepsilon_s] = -\sigma_s^2/(\beta_1 - \alpha_1)$ so neither the demand nor the supply equation satisfies the assumptions of the classical regression model. The price elasticity of demand cannot be consistently estimated by least squares regression of q on x and p. This result is characteristic of simultaneous equations models. Because the endogenous variables are all correlated with the disturbances, the least squares estimators of the parameters of equations with endogenous variables on the right-hand side are inconsistent.[22]

Suppose that we have a sample of T observations on p, q, and x such that

$$\text{plim}(1/T)\mathbf{x}'\mathbf{x} = \sigma_x^2.$$

Because least squares is inconsistent, we might instead use an **instrumental variable estimator**. (See Section 8.3.) The only variable in the system that is not correlated with the disturbances is x. Consider, then, the IV estimator, $\hat{\beta}_1 = (\mathbf{x}'\mathbf{p})^{-1}\mathbf{x}'\mathbf{q}$. This estimator has

$$\text{plim}\,\hat{\beta}_1 = \text{plim}\,\frac{\mathbf{x}'\mathbf{q}/T}{\mathbf{x}'\mathbf{p}/T} = \frac{\sigma_x^2 \beta_1 \alpha_2/(\beta_1 - \alpha_1)}{\sigma_x^2 \alpha_2/(\beta_1 - \alpha_1)} = \beta_1.$$

Evidently, the parameter of the supply curve can be estimated by using an instrumental variable estimator. In the least squares regression of $\mathbf{p}$ on $\mathbf{x}$, the predicted values are $\hat{\mathbf{p}} = (\mathbf{x}'\mathbf{p}/\mathbf{x}'\mathbf{x})\mathbf{x}$. It follows that in the instrumental variable regression the instrument is $\hat{\mathbf{p}}$. That is,

$$\hat{\beta}_1 = \frac{\hat{\mathbf{p}}'\mathbf{q}}{\hat{\mathbf{p}}'\mathbf{p}}.$$

Because $\hat{\mathbf{p}}'\mathbf{p} = \hat{\mathbf{p}}'\hat{\mathbf{p}}$, $\hat{\beta}_1$ is also the slope in a regression of q on these predicted values. This interpretation defines the two-stage least squares estimator.

It would seem natural to use a similar device to estimate the parameters of the demand equation, but unfortunately, we have already used all of the information in the sample. Not only does least squares fail to estimate the demand equation consistently, but without some further assumptions, the sample contains no other information that can be used. This example illustrates the **problem of identification** alluded to in the introduction to this section.

[22]This failure of least squares is sometimes labeled *simultaneous equations bias*.

10.4.2 A GENERAL NOTATION FOR LINEAR SIMULTANEOUS EQUATIONS MODELS[23]

The **structural form** of the model is

$$\begin{aligned}
\gamma_{11}y_{t1} + \gamma_{21}y_{t2} + \cdots + \gamma_{M1}y_{tM} + \beta_{11}x_{t1} + \cdots + \beta_{K1}x_{tK} &= \varepsilon_{t1},\\
\gamma_{12}y_{t1} + \gamma_{22}y_{t2} + \cdots + \gamma_{M2}y_{tM} + \beta_{12}x_{t1} + \cdots + \beta_{K2}x_{tK} &= \varepsilon_{t2},\\
&\vdots\\
\gamma_{1M}y_{t1} + \gamma_{2M}y_{t2} + \cdots + \gamma_{MM}y_{tM} + \beta_{1M}x_{t1} + \cdots + \beta_{KM}x_{tK} &= \varepsilon_{tM}.
\end{aligned} \tag{10-37}$$

There are M equations and M endogenous variables, denoted $y_1, \ldots, y_M$. There are K exogenous variables, $x_1, \ldots, x_K$, that may include predetermined values of $y_1, \ldots, y_M$ as well.[24] The first element of $\mathbf{x}_t$ will usually be the constant, 1. Finally, $\varepsilon_{t1}, \ldots, \varepsilon_{tM}$ are the **structural disturbances**. The subscript t will be used to index observations, $t = 1, \ldots, T$.

In matrix terms, the system may be written

$$[y_1 \quad y_2 \quad \cdots \quad y_M]_t \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1M} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2M} \\ & & \vdots & \\ \gamma_{M1} & \gamma_{M2} & \cdots & \gamma_{MM} \end{bmatrix}$$

$$+ [x_1 \quad x_2 \quad \cdots \quad x_K]_t \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1M} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2M} \\ & & \vdots & \\ \beta_{K1} & \beta_{K2} & \cdots & \beta_{KM} \end{bmatrix} = [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_M]_t,$$

or

$$\mathbf{y}_t'\mathbf{\Gamma} + \mathbf{x}_t'\mathbf{B} = \boldsymbol{\varepsilon}_t'.$$

Each column of the parameter matrices is the vector of coefficients in a particular equation. The underlying theory will imply a number of restrictions on $\mathbf{\Gamma}$ and $\mathbf{B}$. One of the variables in each equation is labeled the *dependent* variable so that its coefficient in the model will be 1. Thus, there will be at least one "1" in each column of $\mathbf{\Gamma}$. This **normalization** is not a substantive restriction. The relationship defined for a given equation will be unchanged if every coefficient in the equation is multiplied by the same constant. Choosing a dependent variable simply removes this indeterminacy. If there are any identities, then the corresponding columns of $\mathbf{\Gamma}$ and $\mathbf{B}$ will be completely known, and there will be no disturbance for that equation. Because not all variables appear in all equations, some of the parameters will be zero. The theory may also impose other types of restrictions on the parameter matrices.

If $\mathbf{\Gamma}$ is an upper triangular matrix, then the system is said to be a **triangular system**. In this case, the model is of the form

[23]We will be restricting our attention to linear models. Nonlinear systems bring forth numerous complications that are beyond the scope of this text. Gallant (1987), Gallant and Holly (1980), Gallant and White (1988), Davidson and MacKinnon (2004), and Wooldridge (2010) provide further discussion.

[24]For the present, it is convenient to ignore the special nature of lagged endogenous variables and treat them the same as strictly exogenous variables.

$$\begin{aligned} y_{t1} &= f_1(\mathbf{x}_t) + \varepsilon_{t1}, \\ y_{t2} &= f_2(y_{t1}, \mathbf{x}_t) + \varepsilon_{t2}, \\ &\vdots \\ y_{tM} &= f_M(y_{t1}, y_{t2}, \ldots, y_{t,M-1}, \mathbf{x}_t) + \varepsilon_{tM}. \end{aligned}$$

The joint determination of the variables is a **recursive model**. The first is completely determined by the exogenous factors. Then, given the first, the second is likewise determined, and so on.

The solution of the system of equations that determines $\mathbf{y}_t$ in terms of $\mathbf{x}_t$ and $\boldsymbol{\varepsilon}_t$ is the reduced form of the model,

$$\begin{aligned} \mathbf{y}_t' &= [x_1 \quad x_2 \quad \cdots \quad x_K]_t \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1M} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2M} \\ & & \vdots & \\ \pi_{K1} & \pi_{K2} & \cdots & \pi_{KM} \end{bmatrix} + [v_1 \quad \cdots \quad v_M]_t \\ &= -\mathbf{x}_t'\mathbf{B}\boldsymbol{\Gamma}^{-1} + \boldsymbol{\varepsilon}_t'\boldsymbol{\Gamma}^{-1} \\ &= \mathbf{x}_t'\boldsymbol{\Pi} + \mathbf{v}_t'. \end{aligned}$$

For this solution to exist, the model must satisfy the **completeness condition** for simultaneous equations systems: $\boldsymbol{\Gamma}$ must be nonsingular.

Example 10.5 Structure and Reduced Form in a Small Macroeconomic Model

Consider the model

$$\begin{aligned} \text{consumption: } c_t &= \alpha_0 + \alpha_1 y_t + \alpha_2 c_{t-1} + \varepsilon_{t,c}, \\ \text{investment: } i_t &= \beta_0 + \beta_1 r_t + \beta_2(y_t - y_{t-1}) + \varepsilon_{t,i}, \\ \text{demand: } y_t &= c_t + i_t + g_t. \end{aligned}$$

The model contains an autoregressive consumption function based on output, y_t, and one lagged value, an investment equation based on interest, r_t, and the growth in output, and an equilibrium condition. The model determines the values of the three endogenous variables c_t, i_t, and y_t. This model is a **dynamic model**. In addition to the exogenous variables r_t and government spending, g_t, it contains two **predetermined variables**, c_{t-1} and y_{t-1}. These are obviously not exogenous, but with regard to the current values of the endogenous variables, they may be regarded as having already been determined. The deciding factor is whether or not they are uncorrelated with the current disturbances, which we might assume. The reduced form of this model is

$$c_t = [\alpha_0(1-\beta_2) + \beta_0\alpha_1 + \alpha_1\beta_1 r_t + \alpha_1 g_t + \alpha_2(1-\beta_2)c_{t-1} - \alpha_1\beta_2 y_{t-1} + (1-\beta_2)\varepsilon_{t,c} + \alpha_1\varepsilon_{t,i}]/\Lambda,$$

$$i_t = [\alpha_0\beta_2 + \beta_0(1-\alpha_1) + \beta_1(1-\alpha_1)r_t + \beta_2 g_t + \alpha_2\beta_2 c_{t-1} - \beta_2(1-\alpha_1)y_{t-1} + \beta_2\varepsilon_{t,c} + (1-\alpha_1)\varepsilon_{t,i}]/\Lambda,$$

$$y_t = [\alpha_0 + \beta_0 + \beta_1 r_t + g_t + \alpha_2 c_{t-1} - \beta_2 y_{t-1} + \varepsilon_{t,c} + \varepsilon_{t,i}]/\Lambda,$$

where $\Lambda = 1 - \alpha_1 - \beta_2$. The completeness condition is that $\alpha_1 + \beta_2$ not equal one. Note that the reduced form preserves the equilibrium condition, $y_t = c_t + i_t + g_t$. Denote $\mathbf{y}' = [c, i, y]$, $\mathbf{x}' = [1, r, g, c_{-1}, y_{-1}]$ and

$$\boldsymbol{\Gamma} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -\alpha_1 & -\beta_2 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -\alpha_0 & -\beta_0 & 0 \\ 0 & -\beta_1 & 0 \\ 0 & 0 & -1 \\ -\alpha_2 & 0 & 0 \\ 0 & \beta_2 & 0 \end{bmatrix}, \quad \boldsymbol{\Gamma}^{-1} = \frac{1}{\Lambda}\begin{bmatrix} 1-\beta_2 & \beta_2 & 1 \\ \alpha_1 & 1-\alpha_1 & 1 \\ \alpha_1 & \beta_2 & 1 \end{bmatrix}.$$

Then, the reduced form coefficient matrix is

$$\boldsymbol{\Pi}' = \frac{1}{\Lambda}\begin{bmatrix} \alpha_0(1-\beta_2)+\beta_0\alpha_1 & \alpha_1\beta_1 & \alpha_1 & \alpha_2(1-\beta_2) & -\beta_2\alpha_1 \\ \alpha_0\beta_2+\beta_0(1-\alpha_1) & \beta_1(1-\alpha_1) & \beta_2 & \alpha_2\beta_2 & -\beta_2(1-\alpha_1) \\ \alpha_0+\beta_0 & \beta_1 & 1 & \alpha_2 & -\beta_2 \end{bmatrix}.$$

There is an ambiguity in the interpretation of coefficients in a simultaneous equations model. The effects in the structural form of the model would be labeled "causal," in that they are derived directly from the underlying theory. However, in order to trace through the effects of autonomous changes in the variables in the model, it is necessary to work through the reduced form. For example, the interest rate does not appear in the consumption function. But that does not imply that changes in r_t would not "cause" changes in consumption, because changes in r_t change investment, which impacts demand which, in turn, does appear in the consumption function. Thus, we can see from the reduced form that $\Delta c_t/\Delta r_t = \alpha_1\beta_1/\Lambda$. Similarly, the "experiment," $\Delta c_t/\Delta y_t$ is meaningless without first determining what caused the change in y_t. If the change were induced by a change in the interest rate, we would find $(\Delta c_t/\Delta r_t)/(\Delta y_t/\Delta r_t) = (\alpha_1\beta_1/\Lambda)/(\beta_1/\Lambda) = \alpha_1$.

The structural disturbances are assumed to be randomly drawn from an M-variate distribution with

$$E[\boldsymbol{\varepsilon}_t|\mathbf{x}_t] = \mathbf{0} \quad \text{and} \quad E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathbf{x}_t] = \boldsymbol{\Sigma}.$$

For the present, we assume that

$$E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_s'|\mathbf{x}_t, \mathbf{x}_s] = \mathbf{0}, \quad \forall t, s.$$

It will occasionally be useful to assume that $\boldsymbol{\varepsilon}_t$ has a multivariate normal distribution, but we shall postpone this assumption until it becomes necessary. It may be convenient to retain the identities without disturbances as separate equations. If so, then one way to proceed with the stochastic specification is to place rows and columns of zeros in the appropriate places in $\boldsymbol{\Sigma}$. It follows that the **reduced-form disturbances**, $\mathbf{v}_t' = \boldsymbol{\varepsilon}_t'\boldsymbol{\Gamma}^{-1}$, have

$$E[\mathbf{v}_t|\mathbf{x}_t] = (\boldsymbol{\Gamma}^{-1})'\mathbf{0} = \mathbf{0},$$
$$E[\mathbf{v}_t\mathbf{v}_t'|\mathbf{x}_t] = (\boldsymbol{\Gamma}^{-1})'\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{-1} = \boldsymbol{\Omega}.$$

This implies that

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}'\boldsymbol{\Omega}\boldsymbol{\Gamma}.$$

The preceding formulation describes the model as it applies to an observation $[\mathbf{y}', \mathbf{x}', \boldsymbol{\varepsilon}']_t$ at a particular point in time or in a cross section. In a sample of data, each joint observation will be one row in a data matrix,

$$[\mathbf{Y} \quad \mathbf{X} \quad \mathbf{E}] = \begin{bmatrix} \mathbf{y}_1' & \mathbf{x}_1' & \boldsymbol{\varepsilon}_1' \\ \mathbf{y}_2' & \mathbf{x}_2' & \boldsymbol{\varepsilon}_2' \\ & \vdots & \\ \mathbf{y}_T' & \mathbf{x}_T' & \boldsymbol{\varepsilon}_T' \end{bmatrix}.$$

In terms of the full set of T observations, the structure is

$$\mathbf{Y\Gamma} + \mathbf{XB} = \mathbf{E},$$

with

$$E[\mathbf{E}|\mathbf{X}] = \mathbf{0} \quad \text{and} \quad E[(1/T)\mathbf{E}'\mathbf{E}|\mathbf{X}] = \boldsymbol{\Sigma}.$$

Under general conditions, we can strengthen this to $\text{plim}[(1/T)\mathbf{E}'\mathbf{E}] = \boldsymbol{\Sigma}$. For convenience in what follows, we will denote a statistic consistently estimating a quantity, such as this one, with

$$(1/T)\mathbf{E}'\mathbf{E} \rightarrow \boldsymbol{\Sigma}.$$

An important assumption is

$$(1/T)\mathbf{X}'\mathbf{X} \rightarrow \mathbf{Q}, \text{ a finite positive definite matrix.} \tag{10-38}$$

We also assume that

$$(1/T)\mathbf{X}'\mathbf{E} \rightarrow \mathbf{0}. \tag{10-39}$$

This assumption is what distinguishes the predetermined variables from the endogenous variables. The reduced form is

$$\mathbf{Y} = \mathbf{X\Pi} + \mathbf{V}, \quad \text{where } \mathbf{V} = \mathbf{E\Gamma}^{-1}.$$

Combining the earlier results, we have

$$\frac{1}{T}\begin{bmatrix} \mathbf{Y}' \\ \mathbf{X}' \\ \mathbf{V}' \end{bmatrix}[\mathbf{Y} \quad \mathbf{X} \quad \mathbf{V}] \rightarrow \begin{bmatrix} \mathbf{\Pi}'\mathbf{Q\Pi} + \mathbf{\Omega} & \mathbf{\Pi}'\mathbf{Q} & \mathbf{\Omega} \\ \mathbf{Q\Pi} & \mathbf{Q} & \mathbf{0}' \\ \mathbf{\Omega} & \mathbf{0} & \mathbf{\Omega} \end{bmatrix}. \tag{10-40}$$

10.4.3 THE IDENTIFICATION PROBLEM

Solving the **identification** problem precedes estimation. We have in hand a certain amount of information to use for inference about the underlying structure consisting of the sample data and theoretical restrictions on the model such as what variables do and do not appear in each of the equations. The issue is whether the information is sufficient to produce estimates of the parameters of the specified model. The case of measurement error that we examined in Section 8.5 is about identification. The sample regression coefficient, b, converges to a function of two underlying parameters, β and σ_u^2; $b = \mathbf{x}'\mathbf{y}/\mathbf{x}'\mathbf{x} \rightarrow \beta/[1 + \sigma_u^2/Q]$, where $(\mathbf{x}'\mathbf{x}/T) \rightarrow Q$. With no further information about σ_u^2, we cannot infer a unique β from the sample information, b and Q—there are different pairs of β and σ_u^2 that are consistent with the same information (b,Q). If there were some nonsample information available, such as $Q = \sigma_u^2$, then there would be a unique solution for β, in particular, $b \rightarrow \beta/2$.

Identification is a theoretical exercise. It arises in all econometric settings in which the parameters of a model are to be deduced from the combination of sample information and nonsample (theoretical) information. The crucial issue is whether it is possible to deduce the values of structural parameters uniquely from the sample information and **nonsample information** provided by theory, mainly restrictions on parameter values. The issue of identification is the subject of a lengthy literature including Working (1927), Bekker and Wansbeek (2001), and continuing through the contemporary discussion of natural experiments [Section 8.8 and Angrist and Pischke (2010), with commentary], instrumental variable estimation in general, and "identification strategies."

The structural model consists of the equation system

$$\mathbf{y}'\mathbf{\Gamma} + \mathbf{x}'\mathbf{B} = \boldsymbol{\varepsilon}'.$$

Each column in $\mathbf{\Gamma}$ and $\mathbf{B}$ are the parameters of a specific equation in the system. The information consists of the sample information, $(\mathbf{Y}, \mathbf{X})$, and other nonsample information in the form of restrictions on parameter matrices. The sample data provide sample moments, $\mathbf{X}'\mathbf{X}/T$, $\mathbf{X}'\mathbf{Y}/T$, and $\mathbf{Y}'\mathbf{Y}/T$. For purposes of identification, suppose we could observe as large a sample as desired. Then, based on our sample information, we could observe [from (10-40)]

$$\begin{aligned}
&(1/T)\mathbf{X}'\mathbf{X} \rightarrow \mathbf{Q},\\
&(1/T)\mathbf{X}'\mathbf{Y} = (1/T)\mathbf{X}'(\mathbf{X\Pi} + \mathbf{V}) \rightarrow \mathbf{Q\Pi},\\
&(1/T)\mathbf{Y}'\mathbf{Y} = (1/T)(\mathbf{X\Pi} + \mathbf{V})'(\mathbf{X\Pi} + \mathbf{V}) \rightarrow \mathbf{\Pi}'\mathbf{Q\Pi} + \mathbf{\Omega}.
\end{aligned}$$

Therefore, $\mathbf{\Pi}$, the matrix of reduced-form coefficients, is observable

$$[(1/T)\mathbf{X}'\mathbf{X}]^{-1}[(1/T)\mathbf{X}'\mathbf{Y}] \rightarrow \mathbf{\Pi}.$$

This estimator is simply the equation-by-equation least squares regression of $\mathbf{Y}$ on $\mathbf{X}$. Because $\mathbf{\Pi}$ is observable, $\mathbf{\Omega}$ is also,

$$[(1/T)\mathbf{Y}'\mathbf{Y}] - [(1/T)\mathbf{Y}'\mathbf{X}][(1/T)\mathbf{X}'\mathbf{X}]^{-1}[(1/T)\mathbf{X}'\mathbf{Y}] \rightarrow \mathbf{\Omega}.$$

This result is the matrix of least squares residual variances and covariances. Therefore,

$\mathbf{\Pi}$ and $\mathbf{\Omega}$ can be estimated consistently by least squares regression of $\mathbf{Y}$ on $\mathbf{X}$.

The information in hand, therefore, consists of $\mathbf{\Pi}$, $\mathbf{\Omega}$, and whatever other nonsample information we have about the structure. The question is whether we can deduce $(\mathbf{\Gamma}, \mathbf{B}, \mathbf{\Sigma})$ from $(\mathbf{\Pi}, \mathbf{\Omega})$. A simple counting exercise immediately reveals that the answer is no—there are M^2 parameters in $\mathbf{\Gamma}$, $M(M + 1)/2$ in $\mathbf{\Sigma}$ and KM in $\mathbf{B}$, to be deduced. The sample data contain KM elements in $\mathbf{\Pi}$ and $M(M + 1)/2$ elements in $\mathbf{\Omega}$. By simply counting equations and unknowns, we find that our data are insufficient by M^2 pieces of information. We have (in principle) used the sample information already, so these M^2 additional restrictions are going to be provided by the theory of the model. The M^2 additional **restrictions** come in the form of normalizations—one coefficient in each equation equals one—most commonly **exclusion restrictions**, which set coefficients to zero and other relationships among the parameters, such as linear relationships, or specific values attached to coefficients. In some instances, restrictions on $\mathbf{\Sigma}$, such as assuming that certain disturbances are uncorrelated, will provide additional information. A small example will help fix ideas.

Example 10.6 Identification of a Supply and Demand Model

Consider a market in which q is quantity of Q, p is price, and z is the price of Z, a related good. We assume that z enters both the supply and demand equations. For example, Z might be a crop that is purchased by consumers and that will be grown by farmers instead of Q if its price rises enough relative to p. Thus, we would expect $\alpha_2 > 0$ and $\beta_2 < 0$. So,

$$\begin{aligned} q_d &= \alpha_0 + \alpha_1 p + \alpha_2 z + \varepsilon_d \qquad (\textit{demand}),\\ q_s &= \beta_0 + \beta_1 p + \beta_2 z + \varepsilon_s \qquad (\textit{supply}),\\ q_d &= q_s = q \qquad\qquad (\textit{equilibrium}). \end{aligned}$$

The reduced form is

$$\begin{aligned} q &= \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1\beta_2 - \alpha_2\beta_1}{\alpha_1 - \beta_1}z + \frac{\alpha_1\varepsilon_s - \beta_1\varepsilon_d}{\alpha_1 - \beta_1} = \pi_{11} + \pi_{21}z + \nu_q,\\ p &= \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2 - \alpha_2}{\alpha_1 - \beta_1}z + \frac{\varepsilon_s - \varepsilon_d}{\alpha_1 - \beta_1} = \pi_{12} + \pi_{22}z + \nu_p. \end{aligned}$$

With only four reduced-form coefficients and six structural parameters, that there will not be a complete solution for all six structural parameters in terms of the four reduced form parameters. This model is unidentified. There is insufficient information in the sample and the theory to deduce the structural parameters.

Suppose, though, that it is known that $\beta_2 = 0$ (farmers do not substitute the alternative crop for this one). Then the solution for β_1 is π_{21}/π_{22}. After a bit of manipulation, we also obtain $\beta_0 = \pi_{11} - \pi_{12}\pi_{21}/\pi_{22}$. The exclusion restriction identifies the supply parameters; $\beta_2 = 0$ excludes z from the supply equation. But this step is as far as we can go. With this restriction, the model becomes partially identified. Some, but not all, of the parameters can be estimated.

Now, suppose that income x, rather than z, appears in the demand equation. The revised model is

$$\begin{aligned} q &= \alpha_0 + \alpha_1 p + \alpha_2 x + \varepsilon_1,\\ q &= \beta_0 + \beta_1 p + \beta_2 z + \varepsilon_2. \end{aligned}$$

Note that one variable is now excluded from each equation. The structure is now

$$[q \quad p]\begin{bmatrix} 1 & 1 \\ -\alpha_1 & -\beta_1 \end{bmatrix} + [1\ x\ z]\begin{bmatrix} -\alpha_0 & -\beta_0 \\ -\alpha_2 & 0 \\ 0 & -\beta_2 \end{bmatrix} = [\varepsilon_1 \quad \varepsilon_2].$$

The reduced form is

$$[q \quad p] = [1\ x\ z]\begin{bmatrix} (\alpha_1\beta_0 - \alpha_0\beta_1)/\Lambda & (\beta_0 - \alpha_0)/\Lambda \\ -\alpha_2\beta_1/\Lambda & -\alpha_2/\Lambda \\ \alpha_1\beta_2/\Lambda & \beta_2/\Lambda \end{bmatrix} + [\nu_1 \quad \nu_2],$$

where $\Lambda = (\alpha_1 - \beta_1)$. The unique solutions for the structural parameters in terms of the reduced-form parameters are now

$$\begin{aligned} \alpha_0 &= \pi_{11} - \pi_{12}\left(\frac{\pi_{31}}{\pi_{32}}\right), & \beta_0 &= \pi_{11} - \pi_{12}\left(\frac{\pi_{21}}{\pi_{22}}\right),\\ \alpha_1 &= \frac{\pi_{31}}{\pi_{32}}, & \beta_1 &= \frac{\pi_{21}}{\pi_{22}},\\ \alpha_2 &= \pi_{22}\left(\frac{\pi_{21}}{\pi_{22}} - \frac{\pi_{31}}{\pi_{32}}\right), & \beta_2 &= \pi_{32}\left(\frac{\pi_{31}}{\pi_{32}} - \frac{\pi_{21}}{\pi_{22}}\right). \end{aligned}$$

With this formulation, all of the parameters are identified. This is an example of an exactly identified model. An additional variation is worth a look. Suppose that a second variable, w (weather), appears in the supply equation,

$$q = \alpha_0 + \alpha_1 p + \alpha_2 x + \varepsilon_1,$$
$$q = \beta_0 + \beta_1 p + \beta_2 z + \beta_3 w + \varepsilon_2.$$

You can easily verify that, the reduced form matrix is the same as the previous one, save for an additional row that contains $[\alpha_1\beta_3/\Lambda, \beta_3/\Lambda]$. This implies that there is now a second solution for α_1, $\pi_{41}/\pi_{42.}$ The two solutions, this and π_{31}/π_{32}, will be different. This model is overidentified. There is more information in the sample and theory than is needed to deduce the structural parameters.

Some equation systems are identified and others are not. The formal mathematical conditions under which an equation in a system is identified turns on two results known as the rank and order conditions. The *order condition* is a simple counting rule. It requires that the number of exogenous variables that appear elsewhere in the equation system must be at least as large as the number of endogenous variables in the equation. (Other specific restrictions on the parameters will be included in this count—note that an "exclusion restriction" is a type of linear restriction.) We used this rule when we constructed the IV estimator in Chapter 8. In that setting, we required our model to be at least *identified* by requiring that the number of instrumental variables not contained in $\mathbf{X}$ be at least as large as the number of endogenous variables. The correspondence of that single equation application with the condition defined here is that the rest of the equation system is the source of the instrumental variables. One simple order condition for identification of an equation system is that each equation contain "its own" exogenous variable that does not appear elsewhere in the system.

The **order condition** is necessary for identification; the **rank condition** is sufficient. The equation system in (10-37) in structural form is $\mathbf{y}'\mathbf{\Gamma} = -\mathbf{x}'\mathbf{B} + \boldsymbol{\varepsilon}'$. The reduced form is $\mathbf{y}' = \mathbf{x}'(-\mathbf{B}\,\mathbf{\Gamma}^{-1}) + \boldsymbol{\varepsilon}'\mathbf{\Gamma}^{-1} = \mathbf{x}'\mathbf{\Pi} + \mathbf{v}'$. The way we are going to deduce the parameters in $(\mathbf{\Gamma}, \mathbf{B}, \mathbf{\Sigma})$ is from the reduced form parameters $(\mathbf{\Pi}, \mathbf{\Omega})$. For the jth equation, the solution is contained in $\mathbf{\Pi\Gamma}_j = -\mathbf{B}_j$, where $\mathbf{\Gamma}_j$ contains all the coefficients in the jth equation that multiply endogenous variables. One of these coefficients will equal one, usually some will equal zero, and the remainder are the nonzero coefficients on endogenous variables in the equation, $\mathbf{Y}_j$ [these are denoted γ_j in (10-41) following]. Likewise, $\mathbf{B}_j$ contains the coefficients in equation j on all exogenous variables in the model—some of these will be zero and the remainder will multiply variables in $\mathbf{X}_j$, the exogenous variables that appear in this equation [these are denoted $\boldsymbol{\beta}_j$ in (10-41) following]. The empirical counterpart will be $\mathbf{Pc}_j = \mathbf{b}_j$, where $\mathbf{P}$ is the estimated reduced form, $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, and $\mathbf{c}_j$ and $\mathbf{b}_j$ will be the estimates of the jth columns of $\mathbf{\Gamma}$ and $\mathbf{B}$. The rank condition ensures that there is a solution to this set of equations. In practical terms, the rank condition is difficult to establish in large equation systems. Practitioners typically take it as a given. In small systems, such as the two-equation systems that dominate contemporary research, it is trivial, as we examine in the next example. We have already used the rank condition in Chapter 8, where it played a role in the relevance condition for instrumental variable estimation. In particular, note after the statement of the assumptions for instrumental variable estimation, we assumed plim$(1/T)\mathbf{Z}'\mathbf{X}$ is a matrix with rank K. (This condition is often labeled the *rank condition* in contemporary applications. It not identical, but it is sufficient for the condition mentioned here.)

Example 10.7 The Rank Condition and a Two-Equation Model

The following two-equation recursive model provides what is arguably the platform for much of contemporary econometric analysis. The main equation of interest is

$$y = \gamma f + \beta x + \varepsilon.$$

The variable f is endogenous (it is correlated with ε.); x is exogenous (it is uncorrelated with ε). The analyst has in hand an instrument for f, z. The instrument, z, is relevant, in that in the auxiliary equation,

$$f = \lambda x + \delta z + w,$$

δ is not zero. The exogeneity assumption is $E[\varepsilon z] = E[wz] = 0$. Note that the source of the endogeneity of f is the assumed correlation of w and ε. For purposes of the exercise, assume that $E[xz] = 0$ and the data satisfy $\mathbf{x'z} = 0$—this actually loses no generality. In this two-equation model, the second equation is already in reduced form; x and z are both exogenous. It follows that λ and δ are estimable by least squares. The estimating equations for (γ, β) are

$$\mathbf{P}\gamma_1 = \begin{bmatrix} \mathbf{x'x} & \mathbf{x'z} \\ \mathbf{z'x} & \mathbf{z'z} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{x'y} & \mathbf{x'f} \\ \mathbf{z'y} & \mathbf{z'f} \end{bmatrix}\begin{pmatrix} 1 \\ -\gamma \end{pmatrix} = \begin{bmatrix} \mathbf{x'y/x'x} & \mathbf{x'f/x'x} \\ \mathbf{z'y/z'z} & \mathbf{z'f/z'z} \end{bmatrix}\begin{pmatrix} 1 \\ -\gamma \end{pmatrix} = \boldsymbol{\beta}_{\text{j}} = \begin{pmatrix} \beta \\ 0 \end{pmatrix}.$$

The solutions are $\gamma = (\mathbf{z'y/z'f})$ and $\beta = (\mathbf{x'y/x'x} - (\mathbf{z'y/z'f})\mathbf{x'f/x'x})$. Because $\mathbf{x'x}$ cannot equal zero, the solution depends on $(\mathbf{z'f/z'z})$ not equal to zero—formally that this part of the reduced form coefficient matrix have rank $M = 1$, which would be the rank condition. Note that the solution for γ is the instrumental variable estimator, with z as instrument for f. (The simplicity of this solution turns on the assumption that $\mathbf{x'z} = 0$. The algebra gets a bit more complicated without it, but the conclusion is the same.)

The rank condition is based on the exclusion restrictions in the model—whether the exclusion restrictions provide enough information to identify the coefficients in the jth equation. Formally, the idea can be developed thusly. With the jth equation written as in (10-41), we call $\mathbf{X}_j$ the *included exogenous variables*. The remaining excluded exogenous variables are denoted $\mathbf{X}_j^*$. The M_j variables $\mathbf{Y}_j$ in (10-41) are the included endogenous variables. With this distinction, we can write the M_j reduced forms for $\mathbf{Y}_j$ as $\overline{\mathbf{\Pi}}_j = \begin{bmatrix} \mathbf{\Pi}_j \\ \mathbf{\Pi}_j^* \end{bmatrix}$. The rank condition (which we state without proof) is that the rank of the lower part of the $M_j \times (K_j + K_j^*)$ matrix, $\overline{\mathbf{\Pi}}_j$, equal M_j. In the preceding example, in the first equation, $\mathbf{Y}_j$ is f, $M_j = 1$, $\mathbf{X}_j$ is x, $\mathbf{X}_j^*$ is z, and $\overline{\mathbf{\Pi}}_j$ is estimated by the regression of f on x and z; $\mathbf{\Pi}_j$ is the coefficient on x and $\mathbf{\Pi}_j^*$ is the coefficient on z. The rank condition we noted earlier is that what is estimated by $\mathbf{z'f/z'z}$, which would correspond to $\mathbf{\Pi}_j^*$ not equal zero, meaning that it has rank 1.

Casual statements of the rank condition based on an IV regression of a variable $\mathbf{y}_{IV}$ on $(M_j + K_j)$ endogenous and exogeneous variables in $\mathbf{X}_{IV}$, using $K_j + K_j^*$ exogenous and instrumental variables in $\mathbf{Z}_{IV}$ (in the most familiar cases, $M_j = K_j^* = 1$), state that the rank requirement is that $(\mathbf{Z}_{\text{IV}}'\mathbf{X}_{IV}/T)$ be nonsingular. In the notation we are using here, $\mathbf{Z}_{IV}$ would be $\mathbf{X} = (\mathbf{X}_j, \mathbf{X}_j^*)$ and $\mathbf{X}_{iv}$ would be $(\mathbf{X}_j, \mathbf{Y}_j)$. This nonsingularity would correspond to full rank of plim$(\mathbf{X'X}/T)$ times plim$[(\mathbf{X'X}^*/T, \mathbf{X'Y}_j/T)]$ because plim$(\mathbf{X'X}/T) = \mathbf{Q}$, which is nonsingular [see (10-40)]. The first K_j columns of this matrix are the last K_j columns of an identity matrix, which have rank K_j. The last M_j columns are estimates of $\mathbf{Q}\overline{\mathbf{\Pi}}_j$, which we require to have rank M_j, so the requirement is that $\overline{\mathbf{\Pi}}_j$ have rank M_j. But, if $K_j^* \geq M_j$ (the order condition), then all that is needed is rank$(\mathbf{\Pi}_j^*) = M_j$, so, in practical terms, the casual statement is correct. It is stronger than necessary; the formal mathematical condition is only that the lower half of the matrix must have rank M_j, but the practical result is much easier to visualize.

It is also easy to verify that the rank condition requires that the predictions of $\mathbf{Y}_j$ using $(\mathbf{X}_j, \mathbf{X}_j^*)\overline{\mathbf{\Pi}}_j$ be linearly independent. Continuing this line of thinking, if we use 2SLS, the rank condition requires that the predicted values of the included endogenous variables not be collinear, which makes sense.

10.4.4 SINGLE EQUATION ESTIMATION AND INFERENCE

For purposes of estimation and inference, we write the model in the way that the researcher would typically formulate it,

$$\begin{aligned}\mathbf{y}_j &= \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{Y}_j\boldsymbol{\gamma}_j + \boldsymbol{\varepsilon}_j \\ &= \mathbf{Z}_j\boldsymbol{\delta}_j + \boldsymbol{\varepsilon}_j,\end{aligned} \tag{10-41}$$

where $\mathbf{y}_j$ is the "dependent variable" in the equation, $\mathbf{X}_j$ is the set of exogenous variables that appear in the jth equation—note that this is not all the variables in the model—and $\mathbf{Z}_j = (\mathbf{X}_j, \mathbf{Y}_j)$. The full set of exogenous variables in the model, including $\mathbf{X}_j$ and variables that appear elsewhere in the model (including a constant term if any equation includes one), is denoted $\mathbf{X}$. For example, in the supply/demand model in Example 10.6, the full set of exogenous variables is $\mathbf{X} = (\mathbf{1}, \mathbf{x}, \mathbf{z})$, while $\mathbf{X}_{Demand} = (\mathbf{1}, \mathbf{x})$ and $\mathbf{X}_{Supply} = (\mathbf{1}, \mathbf{z})$. Finally, $\mathbf{Y}_j$ is the endogenous variables that appear on the right-hand side of the jth equation. Once again, this is likely to be a subset of the endogenous variables in the full model. In Example 10.6, $\mathbf{Y}_j = (\mathbf{price})$ in both cases.

There are two approaches to estimation and inference for simultaneous equations models. **Limited information estimators** are constructed for each equation individually. The approach is analogous to estimation of the seemingly unrelated regressions model in Section 10.2 by least squares, one equation at a time. **Full information estimators** are used to estimate all equations simultaneously. The counterpart for the seemingly unrelated regressions model is the feasible generalized least squares estimator discussed in Section 10.2.3. The major difference to be accommodated at this point is the endogeneity of $\mathbf{Y}_j$ in (10-41).

The equation in (10-41) is precisely the model developed in Chapter 8. Least squares will generally be unsuitable as it is inconsistent due to the correlation between $\mathbf{Y}_j$ and $\boldsymbol{\varepsilon}_j$. The usual approach will be two-stage least squares as developed in Sections 8.3.2 through 8.3.4. The only difference between the case considered here and that in Chapter 8 is the source of the instrumental variables. In our general model in Chapter 8, the source of the instruments remained somewhat ambiguous; the overall rule was "outside the model." In this setting, the instruments come from elsewhere in the model—that is, "not in the jth equation." For estimating the linear simultaneous equations model, the most common estimator is

$$\begin{aligned}\hat{\boldsymbol{\delta}}_{j,\,2\,\mathrm{SLS}} &= [\hat{\mathbf{Z}}_j'\hat{\mathbf{Z}}_j]^{-1}\hat{\mathbf{Z}}_j'\mathbf{y}_j \\ &= [(\mathbf{Z}_j'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}_j)]^{-1}(\mathbf{Z}_j'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_j,\end{aligned} \tag{10-42}$$

where all columns of $\hat{\mathbf{Z}}_j'$ are obtained as predictions in a regression of the corresponding column of $\mathbf{Z}_j$ on $\mathbf{X}$. This equation also results in a useful simplification of the estimated asymptotic covariance matrix,

$$\text{Est.Asy.Var}[\hat{\boldsymbol{\delta}}_{j,\,2\,SLS}] = \hat{\sigma}_{jj}(\hat{\mathbf{Z}}_j'\hat{\mathbf{Z}}_j)^{-1}.$$

It is important to note that σ_{jj} is estimated by

$$\hat{\sigma}_{jj} = \frac{(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_j)'(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_j)}{T}, \tag{10-43}$$

using the original data, not $\hat{\mathbf{Z}}_j$.

Note the role of the order condition for identification in the two-stage least squares estimator. Formally, the order condition requires that the number of exogenous variables that appear elsewhere in the model (not in this equation) be at least as large as the number of endogenous variables that appear in this equation. The implication will be that we are going to predict $\mathbf{Z}_j = (\mathbf{X}_j, \mathbf{Y}_j)$ using $\mathbf{X} = (\mathbf{X}_j, \mathbf{X}_j^*)$. In order for these predictions to be linearly independent, there must be at least as many variables used to compute the predictions as there are variables being predicted. Comparing $(\mathbf{X}_j, \mathbf{Y}_j)$ to $(\mathbf{X}_j, \mathbf{X}_j^*)$, we see that there must be at least as many variables in $\mathbf{X}_j^*$ as there are in $\mathbf{Y}_j$, which is the order condition. The practical rule of thumb that every equation have at least one variable in it that does not appear in any other equation will guarantee this outcome.

Two-stage least squares is used nearly universally in estimation of linear simultaneous equation models—for precisely the reasons outlined in Chapter 8. However, some applications (and some theoretical treatments) have suggested that the **limited information maximum likelihood (LIML) estimator** based on the normal distribution may have better properties. The technique has also found recent use in the analysis of weak instruments. A result that emerges from the derivation is that the LIML estimator has the same asymptotic distribution as the 2SLS estimator, and the latter does not rely on an assumption of normality. This raises the question why one would use the LIML technique given the availability of the more robust (and computationally simpler) alternative. Small sample results are sparse, but they would favor 2SLS as well.[25] One significant virtue of LIML is its invariance to the normalization of the equation. Consider an example in a system of equations,

$$y_1 = y_2\gamma_2 + y_3\gamma_3 + x_1\beta_1 + x_2\beta_2 + \varepsilon_1.$$

An equivalent equation would be

$$\begin{aligned} y_2 &= y_1(1/\gamma_2) + y_3(-\gamma_3/\gamma_2) + x_1(-\beta_1/\gamma_2) + x_2(-\beta_2/\gamma_2) + \varepsilon_1(-1/\gamma_2) \\ &= y_1\widetilde{\gamma}_1 + y_3\widetilde{\gamma}_3 + x_1\widetilde{\beta}_1 + x_2\widetilde{\beta}_2 + \widetilde{\varepsilon}_1. \end{aligned}$$

The parameters of the second equation can be manipulated to produce those of the first. But, as you can easily verify, the 2SLS estimator is not invariant to the normalization of the equation—2SLS would produce numerically different answers. LIML would give the same numerical solutions to both estimation problems suggested earlier. A second virtue is LIML's better performance in the presence of weak instruments.

The LIML, or **least variance ratio** estimator, can be computed as follows.[26] Let

$$\mathbf{W}_j^0 = \mathbf{E}_j^{0\prime}\mathbf{E}_j^0, \tag{10-44}$$

where

$$\mathbf{Y}_j^0 = [\mathbf{y}_j, \mathbf{Y}_j],$$

and

$$\mathbf{E}_j^0 = \mathbf{M}_j\mathbf{Y}_j^0 = [\mathbf{I} - \mathbf{X}_j(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j']\mathbf{Y}_j^0. \tag{10-45}$$

[25]See Phillips (1983).

[26]The LIML estimator was derived by Anderson and Rubin (1949, 1950). [See, also, Johnston (1984).] The much simpler and equally efficient two-stage least squares estimator remains the estimator of choice.

Each column of $\mathbf{E}_j^0$ is a set of least squares residuals in the regression of the corresponding column of $\mathbf{Y}_j^0$ on $\mathbf{X}_j$, that is, only the exogenous variables that appear in the jth equation. Thus, $\mathbf{W}_j^0$ is the matrix of sums of squares and cross products of these residuals. Define

$$\mathbf{W}_j^1 = \mathbf{E}_j^{1\prime}\mathbf{E}_j^1 = \mathbf{Y}_j^{0\prime}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}_j^0. \quad \textbf{(10-46)}$$

That is, $\mathbf{W}_j^1$ is defined like $\mathbf{W}_j^0$ except that the regressions are on all the x's in the model, not just the ones in the jth equation. Let

$$\lambda_1 = \text{smallest characteristic root of } (\mathbf{W}_j^1)^{-1}\mathbf{W}_j^0. \quad \textbf{(10-47)}$$

This matrix is asymmetric, but all its roots are real and greater than or equal to 1. [Depending on the available software, it may be more convenient to obtain the identical smallest root of the symmetric matrix $\mathbf{D} = (\mathbf{W}_j^1)^{-1/2}\mathbf{W}_j^0(\mathbf{W}_j^1)^{-1/2}$.] Now partition $\mathbf{W}_j^0$ into $\mathbf{W}_j^0 = \begin{bmatrix} w_{jj}^0 & \mathbf{w}_j^{0\prime} \\ \mathbf{w}_j^0 & \mathbf{W}_{jj}^0 \end{bmatrix}$ corresponding to $[\mathbf{y}_j, \mathbf{Y}_j]$, and partition $\mathbf{W}_j^1$ likewise. Then, with these parts in hand,

$$\hat{\boldsymbol{\gamma}}_{j,\text{LIML}} = [\mathbf{W}_{jj}^0 - \lambda_1\mathbf{W}_{jj}^1]^{-1}(\mathbf{w}_j^0 - \lambda_1\mathbf{w}_j^1) \quad \textbf{(10-48)}$$

and

$$\hat{\boldsymbol{\beta}}_{j,\text{LIML}} = (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'(\mathbf{y}_j - \mathbf{Y}_j\hat{\boldsymbol{\gamma}}_{j,\text{LIML}}).$$

Note that $\boldsymbol{\beta}_j$ is estimated by a simple least squares regression.[See (3-18).] The asymptotic covariance matrix for the LIML estimator is identical to that for the 2SLS estimator.

Example 10.8 Simultaneity in Health Production

Example 7.1 analyzed the incomes of a subsample of Riphahn, Wambach, and Million's (2003) data on health outcomes in the German Socioeconomic Panel. Here we continue Example 10.4 and consider a Grossman (1972) style model for health and incomes. Our two-equation model is

$$\begin{aligned} \textit{Health Satisfaction} &= \alpha_1 + \gamma_1 \textit{ ln Income} + \alpha_2 \textit{ Female} + \alpha_3 \textit{ Working} + \alpha_4 \textit{ Public} + \alpha_5 \textit{ Add On} \\ &\quad + \alpha_6 \textit{ Age} + \varepsilon_H, \\ \textit{ln Income} &= \beta_1 + \gamma_2 \textit{ Health Satisfaction} + \beta_2 \textit{ Female} + \beta_3 \textit{ Education} + \beta_4 \textit{ Married} \\ &\quad + \beta_5 \textit{HHKids} + \beta_6 \textit{Age} + \varepsilon_I. \end{aligned}$$

For purposes of this application, we avoid panel data considerations by examining only the 1994 wave (cross section) of the data, which contains 3,377 observations. The health outcome variable is *Self Assessed Health Satisfaction* (HSAT). Whether this variable actually corresponds to a commonly defined objective measure of health outcomes is debateable. We will treat it as such. Second, the variable is a scale variable, coded in this data set 0 to 10. [In more recent versions of the GSOEP data, and in the British (BHPS) and Australian (HILDA) counterparts, it is coded 0 to 4.] We would ordinarily treat such a variable as a discrete ordered outcome, as we do in Examples 18.14 and 18.15. We will treat it as if it were continuous in this example, and recognize that there is likely to be some distortion in the measured effects that we are interested in. *Female*, *Working*, *Married*, and *HHkids* are dummy variables, the last indicating whether there are children living in the household. *Education* and *Age* are in years. *Public* and *AddOn* are dummy variables that indicate whether the individual takes up the public health insurance and, if so, whether he or she also takes up the additional

AddOn insurance, which covers some additional costs. Table 10.7 presents OLS and 2SLS estimates of the parameters of the two-equation model. The differences are striking. In the health outcome equation, the OLS coefficient on *ln Income* is quite large (0.42) and highly significant ($t = 5.17$). However, the effect almost doubles in the 2SLS results. The strong negative effect of having the public health insurance might make one wonder if the insurance takeup is endogenous in the same fashion as *ln Income*. (In the original study from which these data were borrowed, the authors were interested in whether takeup of the add on insurance had an impact on usage of the health care system (number of doctor visits). The 2SLS estimates of the *ln Income* equation are also distinctive. Now, the extremely small effect of health estimated by OLS (0.020) becomes the dominant effect, with marital status, in the 2SLS results.

Both equations are overidentified—each has three excluded exogenous variables. Regression of the 2SLS residuals from the HSAT equation on all seven exogenous variables (and the constant) gives an R^2 of 0.0005916, so the chi-squared test of the overidentifying restrictions is $3{,}337(.0005916) = 1.998$. With two degrees of freedom, the critical value is 5.99, so the restrictions would not be rejected. For the *ln Income* equation, the R^2 in the regression of the residuals on all of the exogenous variables is 0.000426, so the test statistic is 1.438, which is not significant. On this basis, we conclude that the specification of the model is adequate.

TABLE 10.7 Estimated Health Production Model (absolute *t* ratios in parentheses)

	Health Equation			*ln Income Equation*		
	OLS	*2SLS*	*LIML*	*OLS*	*2SLS*	*LIML*
Constant	8.903	9.201	9.202	−1.817	−5.379	−5.506
	(40.67)	(30.31)	(30.28)	(30.81)	(8.65)	(8.46)
ln *Income*	0.418	0.710	0.712			
	(5.17)	(3.20)	(3.20)			
Health				0.020	0.497	0.514
				(5.83)	(6.12)	(6.04)
Female	−0.211	−0.218	−0.218	−0.011	0.126	0.131
	(2.76)	(2.85)	(2.85)	(0.70)	(2.78)	(2.79)
Working	0.339	0.259	0.259			
	(3.76)	(2.43)	(2.43)			
Public	−0.472	−0.391	−0.391			
	(4.10)	(3.05)	(3.04)			
Add On	0.204	0.140	0.139			
	(0.80)	(0.54)	(0.54)			
Education				0.055	0.017	0.016
				(17.00)	(1.65)	(1.46)
Married				0.352	0.263	0.260
				(18.11)	(5.08)	(4.86)
Age	−0.038	−0.039	−0.039	−0.002	0.017	0.018
	(11.55)	(11.60)	(11.60)	(2.58)	(4.53)	(4.51)
HHKids				−0.062	−0.061	−0.061
				(3.45)	(1.32)	(1.28)

10.4.5 SYSTEM METHODS OF ESTIMATION

We may formulate the full system of equations as

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \vdots \\ \boldsymbol{\delta}_M \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix} \tag{10-49}$$

or

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon},$$

where

$$E[\boldsymbol{\varepsilon}\,|\,\mathbf{X}] = \mathbf{0}, \quad \text{and} \quad E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\,|\,\mathbf{X}] = \overline{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} \otimes \mathbf{I}. \tag{10-50}$$

[See (10-3).] The least squares estimator,

$$\mathbf{d} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y},$$

is equation-by-equation ordinary least squares and is inconsistent. But even if ordinary least squares were consistent, we know from our results for the seemingly unrelated regressions model that it would be inefficient compared with an estimator that makes use of the cross-equation correlations of the disturbances. For the first issue, we turn once again to an IV estimator. For the second, as we did Section 10.2.1, we use a generalized least squares approach. Thus, assuming that the matrix of instrumental variables, $\overline{\mathbf{W}}$, satisfies the requirements for an IV estimator, a consistent though inefficient estimator would be

$$\hat{\boldsymbol{\delta}}_{\text{IV}} = (\overline{\mathbf{W}}'\mathbf{Z})^{-1}\overline{\mathbf{W}}'\mathbf{y}. \tag{10-51}$$

Analogous to the seemingly unrelated regressions model, a more efficient estimator would be based on the generalized least squares principle,

$$\hat{\boldsymbol{\delta}}_{\text{IV, GLS}} = [\overline{\mathbf{W}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Z}]^{-1}\,\overline{\mathbf{W}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{y}, \tag{10-52}$$

or, where $\mathbf{W}_j$ is the set of instrumental variables for the jth equation,

$$\hat{\boldsymbol{\delta}}_{\text{IV, GLS}} = \begin{bmatrix} \sigma^{11}\mathbf{W}_1'\mathbf{Z}_1 & \sigma^{12}\mathbf{W}_1'\mathbf{Z}_2 & \cdots & \sigma^{1M}W_1'Z_M \\ \sigma^{21}\mathbf{W}_2'\mathbf{Z}_1 & \sigma^{22}\mathbf{W}_2'\mathbf{Z}_2 & \cdots & \sigma^{2M}W_2'Z_M \\ & & \vdots & \\ \sigma^{M1}\mathbf{W}_M'\mathbf{Z}_1 & \sigma^{M2}\mathbf{W}_M'\mathbf{Z}_2 & \cdots & \sigma^{MM}W_M'Z_M \end{bmatrix}^{-1} \begin{bmatrix} \sum_{n=1}^{M}\sigma^{1n}\mathbf{W}_1'\mathbf{y}_n \\ \sum_{n=1}^{M}\sigma^{2n}\mathbf{W}_2'\mathbf{y}_n \\ \vdots \\ \sum_{n=1}^{M}\sigma^{Mn}\mathbf{W}_M'\mathbf{y}_n \end{bmatrix}.$$

Three IV techniques are generally used for joint estimation of the entire system of equations: three-stage least squares, GMM, and **full information maximum likelihood (FIML)**. In the small minority of applications that use a system estimator, 3SLS is usually the estimator of choice. For dynamic models, GMM is sometimes preferred. The FIML estimator is generally of theoretical interest, as it brings no advantage over 3SLS, but is much more complicated to compute.

Consider the IV estimator formed from

$$\overline{\mathbf{W}} = \hat{\mathbf{Z}} = \text{diag}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_1, \ldots, \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_M] = \begin{bmatrix} \hat{\mathbf{Z}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{Z}}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \hat{\mathbf{Z}}_M \end{bmatrix}.$$

The IV estimator,

$$\hat{\boldsymbol{\delta}}_{\text{IV}} = [\hat{\mathbf{Z}}'\mathbf{Z}]^{-1}\hat{\mathbf{Z}}'\mathbf{y},$$

is simply equation-by-equation 2SLS. We have already established the consistency of 2SLS. By analogy to the seemingly unrelated regressions model of Section 10.2, however, we would expect this estimator to be less efficient than a GLS estimator. A natural candidate would be

$$\hat{\boldsymbol{\delta}}_{\text{3SLS}} = [\hat{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Z}]^{-1}\hat{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{y}.$$

For this estimator to be a valid IV estimator, we must establish that

$$\text{plim}\, \frac{1}{T}\hat{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\boldsymbol{\varepsilon} = \mathbf{0},$$

which is M sets of equations, each one of the form

$$\text{plim}\, \frac{1}{T}\sum_{j=1}^{M}\sigma^{ij}\hat{\mathbf{Z}}_i'\boldsymbol{\varepsilon}_j = \mathbf{0}.$$

Each is the sum of vectors, all of which converge to zero, as we saw in the development of the 2SLS estimator. The second requirement, that

$$\text{plim}\, \frac{1}{T}\hat{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Z} \neq \mathbf{0},$$

and that the matrix be nonsingular, can be established along the lines of its counterpart for 2SLS. Identification of every equation by the rank condition is sufficient.

Once again, using the idempotency of $\mathbf{I} - \mathbf{M}$, we may also interpret this estimator as a GLS estimator of the form

$$\hat{\boldsymbol{\delta}}_{\text{3SLS}} = [\hat{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\hat{\mathbf{Z}})^{-1}\hat{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})]\mathbf{y}. \quad \textbf{(10-53)}$$

The appropriate asymptotic covariance matrix for the estimator is

$$\text{Asy.Var}[\hat{\boldsymbol{\delta}}_{\text{3SLS}}] = (\overline{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\overline{\mathbf{Z}})^{-1}, \quad \textbf{(10-54)}$$

where $\overline{\mathbf{Z}} = \text{diag}[\mathbf{X}\boldsymbol{\Pi}_j, \mathbf{X}_j]$. This matrix would be estimated with the bracketed inverse matrix in (10-53).

Using sample data, we find that $\overline{\mathbf{Z}}$ may be estimated with $\hat{\mathbf{Z}}$. The remaining difficulty is to obtain an estimate of $\boldsymbol{\Sigma}$. In estimation of the seemingly unrelated regressions model, for efficient estimation, any consistent estimator of $\boldsymbol{\Sigma}$ will do. The designers of the 3SLS method, Zellner and Theil (1962), suggest the natural choice arising out of the two-stage least estimates. The **three-stage least squares (3SLS) estimator** is thus defined as follows:

1. Estimate $\boldsymbol{\Pi}$ by ordinary least squares and compute $\hat{\mathbf{Y}}_m$ for each equation.
2. Compute $\hat{\boldsymbol{\delta}}_{m,\,\text{2SLS}}$ for each equation; then

$$\hat{\sigma}_{mn} = \frac{(\mathbf{y}_m - \mathbf{Z}_m\hat{\hat{\boldsymbol{\delta}}}_m)'(\mathbf{y}_n - \mathbf{Z}_n\hat{\hat{\boldsymbol{\delta}}}_n)}{T}. \tag{10-55}$$

3. Compute the GLS estimator according to (10-53) and an estimate of the asymptotic covariance matrix according to (10-54) using $\hat{\mathbf{Z}}$ and $\hat{\boldsymbol{\Sigma}}$.

By showing that the 3SLS estimator satisfies the requirements for an IV estimator, we have established its consistency. The question of asymptotic efficiency remains. It can be shown that of all IV estimators that use only the sample information embodied in the system, 3SLS is asymptotically efficient.

Example 10.9 Klein's Model I

A widely used example of a simultaneous equations model of the economy is Klein's (1950) Model I. The model may be written

$$\begin{aligned}
C_t &= \alpha_0 + \alpha_1 P_t + \alpha_2 P_{t-1} + \alpha_3(W_t^p + W_t^g) + \varepsilon_{1t} && \text{(consumption)},\\
I_t &= \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + \beta_3 K_{t-1} + \varepsilon_{2t} && \text{(investment)},\\
W_t^p &= \gamma_0 + \gamma_1 X_t + \gamma_2 X_{t-1} + \gamma_3 A_t + \varepsilon_{3t} && \text{(private wages)},\\
X_t &= C_t + I_t + G_t && \text{(equilibrium demand)},\\
P_t &= X_t - T_t - W_t^p && \text{(private profits)},\\
K_t &= K_{t-1} + I_t && \text{(capital stock)}.
\end{aligned}$$

The endogenous variables are each on the left-hand side of an equation and are labeled on the right. The exogenous variables are G_t = government nonwage spending, T_t = indirect business taxes plus net exports, W_t^g = government wage bill, A_t = time trend measured as years from 1931, and the constant term. There are also three predetermined variables: the lagged values of the capital stock, private profits, and total demand. The model contains three **behavioral equations**, an **equilibrium condition**, and two accounting identities. This model provides an excellent example of a small, dynamic model of the economy. It has also been widely used as a test ground for simultaneous equations estimators. Klein estimated the parameters using yearly aggregate data for the U.S. for 1921 to 1941. The data are listed in Appendix Table F10.3. Table 10.8 presents limited and full information estimates for Klein's Model I based on the original data.

It might seem, in light of the entire discussion, that one of the structural estimators described previously should always be preferred to ordinary least squares, which alone among the estimators considered here is inconsistent. Unfortunately, the issue is not so clear. First, it is often found that the OLS estimator is surprisingly close to the structural estimator. It can be shown that, at least in some cases, OLS has a smaller variance about its mean than does 2SLS about its mean, leading to the possibility that OLS might be more precise in a mean-squared-error sense. But this result must be tempered by the finding that the OLS standard errors are, in all likelihood, not useful for inference purposes. Obviously, this discussion is relevant only to finite samples. Asymptotically, 2SLS must dominate OLS, and in a correctly specified model, any full information estimator (3SLS) must dominate any limited information one (2SLS). The finite sample properties are of crucial importance. Most of what we know is asymptotic properties, but most applications are based on rather small or moderately sized samples.

Although the system methods of estimation are asymptotically better, they have two problems. First, any specification error in the structure of the model will be propagated throughout the system by 3SLS. The limited information estimators will, by and large,

TABLE 10.8 Estimates of Klein's Model I (Estimated asymptotic standard errors in parentheses)

	2SLS				**3SLS**			
C	16.6	0.017	0.216	0.810	16.4	0.125	0.163	0.790
	(1.32)	(0.118)	(0.107)	(0.040)	(1.30)	(0.108)	(0.100)	(0.038)
I	20.3	0.150	0.616	−0.158	28.2	−0.013	0.756	−0.195
	(7.54)	(0.173)	(0.162)	(0.036)	(6.79)	(0.162)	(0.153)	(0.033)
W^P	1.50	0.439	0.147	0.130	1.80	0.400	0.181	0.150
	(1.15)	(0.036)	(0.039)	(0.029)	(1.12)	(0.032)	(0.034)	(0.028)
	LIML				**OLS**			
C	17.1	−0.222	0.396	0.823	16.2	0.193	0.090	0.796
	(1.84)	(0.202)	(0.174)	(0.055)	(1.30)	(0.091)	(0.091)	(0.040)
I	22.6	0.075	0.680	−0.168	10.1	0.480	0.333	−0.112
	(9.24)	(0.219)	(0.203)	(0.044)	(5.47)	(0.097)	(0.101)	(0.027)
W^P	1.53	0.434	0.151	0.132	1.50	0.439	0.146	0.130
	(2.40)	(0.137)	(0.135)	(0.065)	(1.27)	(0.032)	(0.037)	(0.032)

confine a problem to the particular equation in which it appears. Second, in the same fashion as the SUR model, the finite-sample variation of the estimated covariance matrix is transmitted throughout the system. Thus, the finite-sample variance of 3SLS may well be as large as or larger than that of 2SLS.[27]

10.5 SUMMARY AND CONCLUSIONS

This chapter has surveyed the specification and estimation of multiple equations models. The SUR model is an application of the generalized regression model introduced in Chapter 9. The advantage of the SUR formulation is the rich variety of behavioral models that fit into this framework. We began with estimation and inference with the SUR model, treating it essentially as a generalized regression. The major difference between this set of results and the single-equation model in Chapter 9 is practical. While the SUR model is, in principle, a single equation GR model with an elaborate covariance structure, special problems arise when we explicitly recognize its intrinsic nature as a set of equations linked by their disturbances. The major result for estimation at this step is the feasible GLS estimator. In spite of its apparent complexity, we can estimate the SUR model by a straightforward two-step GLS approach that is similar to the one we used for models with heteroscedasticity in Chapter 9. We also extended the SUR model to autocorrelation and heteroscedasticity. Once again, the multiple equation nature of the model complicates these applications. Section 10.4 presented a common application of the seemingly unrelated regressions model, the estimation of demand systems. One of the signature features of this literature is the seamless transition from the theoretical models of optimization of consumers and producers to the sets of empirical demand equations derived from Roy's identity for consumers and Shephard's lemma for producers.

[27]See Cragg (1967) and the many related studies listed by Judge et al. (1985, pp. 646–653).

The multiple equations models surveyed in this chapter involve most of the issues that arise in analysis of linear equations in econometrics. Before one embarks on the process of estimation, it is necessary to establish that the sample data actually contain sufficient information to provide estimates of the parameters in question. This is the question of identification. Identification involves both the statistical properties of estimators and the role of theory in the specification of the model. Once identification is established, there are numerous methods of estimation. We considered three single-equation techniques, least squares, instrumental variables, and maximum likelihood. Fully efficient use of the sample data will require joint estimation of all the equations in the system. Once again, there are several techniques-these are extensions of the single-equation methods including three-stage least squares-and full information maximum likelihood. In both frameworks, this is one of those benign situations in which the computationally simplest estimator is generally the most efficient one.

Key Terms and Concepts

- Behavioral equation
- Cobb–Douglas model
- Complete system of equations
- Completeness condition
- Constant returns to scale
- Demand system
- Dynamic model
- Econometric model
- Equilibrium condition
- Exclusion restrictions
- Exogenous
- Flexible functional
- Full information estimator
- Full information maximum likelihood (FIML)
- Generalized regression model
- Homogeneity restriction
- Identical explanatory variables
- Identification
- Instrumental variable estimator
- Interdependent
- Invariance
- Jointly dependent
- Kronecker product
- Least variance ratio
- Likelihood ratio test
- Limited information estimator
- Limited information maximum likelihood (LIML) estimator
- Nonsample information
- Normalization
- Order condition
- Pooled model
- Predetermined variable
- Problem of identification
- Rank condition
- Reduced form
- Reduced-form disturbance
- Restrictions
- Seemingly unrelated regressions (SUR)
- Share equations
- Shephard's lemma
- Singular disturbance covariance matrix
- Simultaneous equations bias
- Structural disturbance
- Structural equation
- Structural form
- Systems of demand equations
- Three-stage least squares (3SLS) estimator
- Translog function
- Triangular system

Exercises

1. A sample of 100 observations produces the following sample data:

$$\bar{y}_1 = 1, \bar{y}_2 = 2, \mathbf{y}_1'\mathbf{y}_1 = 150, \mathbf{y}_2'\mathbf{y}_2 = 550, \mathbf{y}_1'\mathbf{y}_2 = 260.$$

The underlying seemingly unrelated regressions model is

$$y_1 = \mu + \varepsilon_1,$$
$$y_2 = \mu + \varepsilon_2.$$

a. Compute the OLS estimate of μ, and estimate the sampling variance of this estimator.

b. Compute the FGLS estimate of μ and the sampling variance of the estimator.

2. Consider estimation of the following two-equation model:

$$y_1 = \beta_1 + \varepsilon_1,$$
$$y_2 = \beta_2 x + \varepsilon_2.$$

A sample of 50 observations produces the following moment matrix:

$$\begin{array}{c} \\ 1 \\ y_1 \\ y_2 \\ x \end{array} \begin{array}{c} \begin{array}{cccc} 1 & y_1 & y_2 & x \end{array} \\ \begin{bmatrix} 50 & & & \\ 150 & 500 & & \\ 50 & 40 & 90 & \\ 100 & 60 & 50 & 100 \end{bmatrix} \end{array}.$$

a. Write the explicit formula for the GLS estimator of $[\beta_1, \beta_2]$. What is the asymptotic covariance matrix of the estimator?

b. Derive the OLS estimator and its sampling variance in this model.

c. Obtain the OLS estimates of β_1 and β_2, and estimate the sampling covariance matrix of the two estimates. Use n instead of $(n - 1)$ as the divisor to compute the estimates of the disturbance variances.

d. Compute the FGLS estimates of β_1 and β_2 and the estimated sampling covariance matrix.

e. Test the hypothesis that $\beta_2 = 1$.

3. The model

$$y_1 = \beta_1 x_1 + \varepsilon_1,$$
$$y_2 = \beta_2 x_2 + \varepsilon_2$$

satisfies all the assumptions of the seemingly unrelated regressions model. All variables have zero means. The following sample second-moment matrix is obtained from a sample of 20 observations:

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

a. Compute the FGLS estimates of β_1 and β_2.

b. Test the hypothesis that $\beta_1 = \beta_2$.

c. Compute the maximum likelihood estimates of the model parameters.

d. Use the likelihood ratio test to test the hypothesis in part b.

4. Prove that in the model

$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1,$$
$$\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2,$$

generalized least squares is equivalent to equation-by-equation ordinary least squares if $\mathbf{X}_1 = \mathbf{X}_2$. The general case is considered in Exercise 14.

5. Consider the two-equation system

$$y_1 = \beta_1 x_1 + \varepsilon_1,$$
$$y_2 = \beta_2 x_2 + \beta_3 x_3 + \varepsilon_2.$$

Assume that the disturbance variances and covariance are known. Now suppose that the analyst of this model applies GLS but erroneously omits x_3 from the second equation. What effect does this specification error have on the consistency of the estimator of β_1?

6. Consider the system

$$y_1 = \alpha_1 + \beta x + \varepsilon_1,$$
$$y_2 = \alpha_2 + \varepsilon_2.$$

The disturbances are freely correlated. Prove that GLS applied to the system leads to the OLS estimates of α_1 and α_2 but to a mixture of the least squares slopes in the regressions of y_1 and y_2 on x as the estimator of β. What is the mixture? To simplify the algebra, assume (with no loss of generality) that $\bar{x} = 0$.

7. For the model

$$y_1 = \alpha_1 + \beta x + \varepsilon_1,$$
$$y_2 = \alpha_2 + \varepsilon_2,$$
$$y_3 = \alpha_3 + \varepsilon_3,$$

assume that $y_{i2} + y_{i3} = 1$ at every observation. Prove that the sample covariance matrix of the least squares residuals from the three equations will be singular, thereby precluding computation of the FGLS estimator. How could you proceed in this case?

8. Consider the following two-equation model:

$$y_1 = \gamma_1 y_2 + \beta_{11} x_1 + \beta_{21} x_2 + \beta_{31} x_3 + \varepsilon_1,$$
$$y_2 = \gamma_2 y_1 + \beta_{12} x_1 + \beta_{22} x_2 + \beta_{32} x_3 + \varepsilon_2.$$

a. Verify that, as stated, neither equation is identified.
b. Establish whether or not the following restrictions are sufficient to identify (or partially identify) the model:
   (1) $\beta_{21} = \beta_{32} = 0,$
   (2) $\beta_{12} = \beta_{22} = 0,$
   (3) $\gamma_1 = 0,$
   (4) $\gamma_1 = \gamma_2$ and $\beta_{32} = 0,$
   (5) $\sigma_{12} = 0$ and $\beta_{31} = 0,$
   (6) $\gamma_1 = 0$ and $\sigma_{12} = 0,$
   (7) $\beta_{21} + \beta_{22} = 1,$
   (8) $\sigma_{12} = 0, \beta_{21} = \beta_{22} = \beta_{31} = \beta_{32} = 0,$
   (9) $\sigma_{12} = 0, \beta_{11} = \beta_{21} = \beta_{22} = \beta_{31} = \beta_{32} = 0.$

9. Obtain the reduced form for the model in Exercise 8 under each of the assumptions made in parts a and in parts b(1) and b(9).
10. The following model is specified:

$$y_1 = \gamma_1 y_2 + \beta_{11}x_1 + \varepsilon_1,$$
$$y_2 = \gamma_2 y_1 + \beta_{22}x_2 + \beta_{32}x_3 + \varepsilon_2.$$

All variables are measured as deviations from their means. The sample of 25 observations produces the following matrix of sums of squares and cross products:

$$\begin{array}{c} \\ y_1 \\ y_2 \\ x_1 \\ x_2 \\ x_3 \end{array}\begin{array}{c} \begin{array}{ccccc} y_1 & y_2 & x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} 20 & 6 & 4 & 3 & 5 \\ 6 & 10 & 3 & 6 & 7 \\ 4 & 3 & 5 & 2 & 3 \\ 3 & 6 & 2 & 10 & 8 \\ 5 & 7 & 3 & 8 & 15 \end{bmatrix} \end{array}.$$

a. Estimate the two equations by OLS.
b. Estimate the parameters of the two equations by 2SLS. Also estimate the asymptotic covariance matrix of the 2SLS estimates.
c. Obtain the LIML estimates of the parameters of the first equation.
d. Estimate the two equations by 3SLS.
e. Estimate the reduced form coefficient matrix by OLS and indirectly by using your structural estimates from part b.

11. For the model

$$y_1 = \gamma_1 y_2 + \beta_{11}x_1 + \beta_{21}x_2 + \varepsilon_1,$$
$$y_2 = \gamma_2 y_1 + \beta_{32}x_3 + \beta_{42}x_4 + \varepsilon_2$$

show that there are two restrictions on the reduced form coefficients. Describe a procedure for estimating the model while incorporating the restrictions.

12. Prove that

$$\text{plim}\, \frac{\mathbf{Y}_m'\boldsymbol{\varepsilon}_m}{T} = \underline{\boldsymbol{\omega}}_m - \boldsymbol{\Omega}_{mm}\boldsymbol{\gamma}_m.$$

13. Prove that an underidentified equation cannot be estimated by 2SLS.
14. Prove the general result in point 2 in Section 10.2.2, if the $\mathbf{X}$ matrices in (10-1) are identical, then full GLS is equation-by-equation OLS. *Hints:* If all the $\mathbf{X}_m$ matrices are identical, then the inverse matrix in (10-10) is $[\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}'\mathbf{X}]^{-1}$. Also, $\mathbf{X}_m'\mathbf{y}_m = \mathbf{X}'\mathbf{y}_m = \mathbf{X}'\mathbf{X}\mathbf{b}_m$. Use these results to show that for the first equation,

$$\hat{\boldsymbol{\beta}}_1 = \sum_{n=1}^{M}\sigma_{1n}\sum_{l=1}^{M}\sigma^{nl}\mathbf{b}_l = \mathbf{b}_1\left(\sum_{n=1}^{M}\sigma_{1n}\sigma^{n1}\right) + \mathbf{b}_2\left(\sum_{n=1}^{M}\sigma_{1n}\sigma^{n2}\right) + \cdots$$
$$+ \mathbf{b}_M\left(\sum_{n=1}^{M}\sigma_{1n}\sigma^{nM}\right),$$

and likewise for the others.

Applications

Some of these applications will require econometric software for the computations. The calculations are standard, and are available as commands in, for example, *Stata, SAS, E-Views* or *LIMDEP*, or as existing programs in *R*.

1. **Statewide aggregate production function.** Continuing Example 10.1, data on output, the capital stocks, and employment are aggregated 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

$$w_{nt} = emp_{nt} / \sum_{j=1}^{M_m} emp_{jt},$$

where M_m is the number of states in region m. Then, the unemployment rate for region m at time t is the following average of the unemployment rates of the states (n) in region (m) at time t:

$$unemp_{mt} = \Sigma_j w_{nt}(j) unemp_{nt}(j).$$

2. Continuing the analysis of Section 10.3.2, we find that a translog cost function for one output and three factor inputs that does not impose constant returns to scale is

$$\begin{aligned} \ln C = \alpha &+ \beta_1 \ln p_1 + \beta_2 \ln p_2 + \beta_3 \ln p_3 + \delta_{11} \tfrac{1}{2} \ln^2 p_1 + \delta_{12} \ln p_1 \ln p_2 \\ &+ \delta_{13} \ln p_1 \ln p_3 + \delta_{22} \tfrac{1}{2} \ln^2 p_2 + \delta_{23} \ln p_2 \ln p_3 + \delta_{33} \tfrac{1}{2} \ln^2 p_3 \\ &+ \gamma_{q1} \ln Q \ln p_1 + \gamma_{q2} \ln Q \ln p_2 + \gamma_{q3} \ln Q \ln p_3 \\ &+ \beta_q \ln Q + \beta_{qq} \tfrac{1}{2} \ln^2 Q + \varepsilon_c. \end{aligned}$$

The factor share equations are

$$\begin{aligned} S_1 &= \beta_1 + \delta_{11} \ln p_1 + \delta_{12} \ln p_2 + \delta_{13} \ln p_3 + \gamma_{q1} \ln Q + \varepsilon_1, \\ S_2 &= \beta_2 + \delta_{12} \ln p_1 + \delta_{22} \ln p_2 + \delta_{23} \ln p_3 + \gamma_{q2} \ln Q + \varepsilon_2, \\ S_3 &= \beta_3 + \delta_{13} \ln p_1 + \delta_{23} \ln p_2 + \delta_{33} \ln p_3 + \gamma_{q3} \ln Q + \varepsilon_3. \end{aligned}$$

[See Christensen and Greene (1976) for analysis of this model.]

a. The three factor shares must add identically to 1. What restrictions does this requirement place on the model parameters?
b. Show how the adding-up condition in (10-33) can be imposed directly on the model by specifying the translog model in (C/p_3), (p_1/p_3), and (p_2/p_3) and dropping the third share equation. (See Example 10.3.) Notice that this reduces the number of free parameters in the model to 10.
c. Continuing part b, the model as specified with the symmetry and equality restrictions has 15 parameters. By imposing the constraints, you reduce this number to 10 in the estimating equations. How would you obtain estimates of the parameters not estimated directly?
d. Estimate each of the three equations you obtained in part b by ordinary least squares. Do the estimates appear to satisfy the cross-equation equality and symmetry restrictions implied by the theory?

e. Using the data in Section 10.3.1, estimate the full system of three equations (cost and the two independent shares), imposing the symmetry and cross-equation equality constraints.

f. Using your parameter estimates, compute the estimates of the elasticities in (10-34) at the means of the variables.

g. Use a likelihood ratio statistic to test the joint hypothesis that $\gamma_{qi} = 0, i = 1, 2, 3.$ [*Hint:* Just drop the relevant variables from the model.]

3. The Grunfeld investment data in Appendix Table 10.4 constitute a classic data set that has been used for decades to develop and demonstrate estimators for seemingly unrelated regressions.[28] Although somewhat dated at this juncture, they remain an ideal application of the techniques presented in this chapter. The data consist of time series of 20 yearly observations on 10 firms. The three variables are

$$I_{it} = \text{gross investment},$$
$$F_{it} = \text{market value of the firm at the end of the previous year},$$
$$C_{it} = \text{value of the stock of plant and equipment at the end of the previous year.}$$

The main equation in the studies noted is

$$I_{it} = \beta_1 + \beta_2 F_{it} + \beta_3 C_{it} + \varepsilon_{it}.$$

a. Fit the 10 equations separately by ordinary least squares and report your results.

b. Use a Wald (Chow) test to test the "aggregation" restriction that the 10 coefficient vectors are the same.

c. Use the seemingly unrelated regressions (FGLS) estimator to reestimate the parameters of the model, once again, allowing the coefficients to differ across the 10 equations. Now, use the pooled model and, again, FGLS, to estimate the constrained equation with equal parameter vectors, and test the aggregation hypothesis.

d. Using the OLS residuals from the separate regressions, use the LM statistic in (10-17) to test for the presence of cross-equation correlation.

e. An alternative specification to the model in part c that focuses on the variances rather than the means is a groupwise heteroscedasticity model. For the current application, you can fit this model using (10-20), (10-21), and (10-22), while imposing the much simpler model with $\sigma_{ij} = 0$ when $i \neq j$. Do the results of the pooled model differ in the two cases considered, simple OLS and groupwise heteroscedasticity?

4. The data in Appendix Table F5.2 may be used to estimate a small macroeconomic model. Use these data to estimate the model in Example 10.5. Estimate the parameters of the two equations by two-stage and three-stage least squares.

5. Using the cost function estimates in Example 10.2, we obtained an estimate of the efficient scale, $Q* = \exp[(1 - \beta_q)/(2\beta_{qq})]$. We can use the delta method in Section 4.5.4 to compute an asymptotic standard error for the estimator of Q^* and a confidence interval. The estimators of the two parameters are $b_q = 0.23860$ and $b_{qq} = 0.04506$. The estimates of the asymptotic covariance matrix are

[28]See Grunfeld (1958), Grunfeld and Griliches (1960), Boot and de Witt (1960), and Kleiber and Zeileis (2010).

$v_q = 0.00344554$, $v_{qq} = 0.0000258021$, $c_{q,qq} = -0.000291067$. Use these results to form a 95% confidence interval for Q^*. (*Hint:* $\partial Q*/\partial b_j = Q*\partial \ln Q*/\partial b_j$.)

6. Using the estimated health outcomes model in Example 10.8, determine the expected values of *ln Income* and *Health Satisfaction* for a person with the following characteristics: $Female = 1$, $Working = 1$, $Public = 1$, $AddOn = 0$, $Education = 14$, $Married = 1$, $HHKids = 1$, $Age = 35$. Now, repeat the calculation with the same person but with $Age = 36$. Likewise, with $Female = 0$ (and $Age = 35$). Note, the sample range of *Income* is 0 – 3.0, with sample mean approximately 0.4. The income data are in 10,000DM units (pre-Euro). In both cases, note how the health satisfaction outcome changes when the exogenous variable (*Age* or *Female*) changes (by one unit).