15
SIMULTANEOUS-EQUATIONS MODELS

15.1 INTRODUCTION

Although most of our work thus far has been in the context of single-equation models, even a cursory look through almost any economics textbook shows that much of the theory is built on sets, or *systems,* of relationships. Familiar examples include market equilibrium, models of the macroeconomy, and sets of factor or commodity demand equations. Whether one's interest is only in a particular part of the system or in the system as a whole, the interaction of the variables in the model will have important implications for both interpretation and estimation of the model's parameters. The implications of simultaneity for econometric estimation were recognized long before the apparatus discussed in this chapter was developed.[1] The subsequent research in the subject, continuing to the present, is among the most extensive in econometrics.

This chapter considers the issues that arise in interpreting and estimating multiple-equations models. Section 15.2 describes the general framework used for analyzing systems of simultaneous equations. Most of the discussion of these models centers on problems of estimation. But before estimation can even be considered, the fundamental question of whether the parameters of interest in the model are even estimable must be resolved. This **problem of identification** is discussed in Section 15.3. Sections 15.4 to 15.7 then discuss methods of estimation. Section 15.8 is concerned with specification tests. In Section 15.9, the special characteristics of dynamic models are examined.

15.2 FUNDAMENTAL ISSUES IN SIMULTANEOUS-EQUATIONS MODELS

In this section, we describe the basic terminology and statistical issues in the analysis of simultaneous-equations models. We begin with some simple examples and then present a general framework.

15.2.1 ILLUSTRATIVE SYSTEMS OF EQUATIONS

A familiar example of a system of simultaneous equations is a model of market equilibrium, consisting of the following:

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

[1] See, for example, Working (1926) and Haavelmo (1943).

These equations are **structural equations** in that they are derived from theory and each purports to describe a particular aspect of the economy.[2] Since 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, since 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.** *The completeness of the system requires that the number of equations equal the number of endogenous variables.* As a general rule, it is not possible to estimate all the parameters of incomplete systems (although it may be possible to estimate some of them).

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

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

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, and ε_d, and ε_s produces the **reduced form** of the model

$$\begin{aligned} p &= \frac{\alpha_2 x}{\beta_1 - \alpha_1} + \frac{\varepsilon_d - \varepsilon_s}{\beta_1 - \alpha_1} = \pi_1 x + v_1, \\ q &= \frac{\beta_1 \alpha_2 x}{\beta_1 - \alpha_1} + \frac{\beta_1 \varepsilon_d - \alpha_1 \varepsilon_s}{\beta_1 - \alpha_1} = \pi_2 x + v_2. \end{aligned} \tag{15-1}$$

(Note the role of the "completeness" requirement that α_1 not equal β_1.)

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 y 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.[3]

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

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

Since least squares is inconsistent, we might instead use an **instrumental variable estimator.**[4] The only variable in the system that is not correlated with the disturbances is x.

[2] The distinction between **structural** and **nonstructural** models is sometimes drawn on this basis. See, for example, Cooley and LeRoy (1985).

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

[4] See Section 5.4.

Consider, then, the IV estimator, $\hat{\beta}_1 = \mathbf{q}'\mathbf{x}/\mathbf{p}'\mathbf{x}$. This estimator has

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

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 $\hat{\mathbf{p}} = (\mathbf{p}'\mathbf{x}/\mathbf{x}'\mathbf{x})\mathbf{x}$. It follows that in the instrumental variable regression the instrument is $\hat{\mathbf{p}}$. That is,

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

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

It would be desirable to use a similar device to estimate the parameters of the demand equation, but unfortunately, we have exhausted the information in the sample. Not only does least squares fail to estimate the demand equation, 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 chapter.

A second example is the following simple model of income determination.

Example 15.1 A Small Macroeconomic Model

Consider the model,

$$\begin{aligned} &\textit{consumption:} & c_t &= \alpha_0 + \alpha_1 y_t + \alpha_2 c_{t-1} + \varepsilon_{t1}, \\ &\textit{investment:} & i_t &= \beta_0 + \beta_1 r_t + \beta_2(y_t - y_{t-1}) + \varepsilon_{t2}, \\ &\textit{demand:} & y_t &= c_t + i_t + g_t. \end{aligned}$$

The model contains an autoregressive consumption function, an investment equation based on interest 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 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

$$\begin{aligned} Ac_t &= \alpha_0(1 - \beta_2) + \beta_0\alpha_1 + \alpha_1\beta_1 r_t + \alpha_1 g_t + \alpha_2(1 - \beta_2)c_{t-1} - \alpha_1\beta_2 y_{t-1} + (1 - \beta_2)\varepsilon_{t1} + \alpha_1\varepsilon_{t2}, \\ Ai_t &= \alpha_0\beta_2 + \beta_0(1 - \alpha_1) + \beta_1(1 - \alpha_1)r_t + \beta_2 g_t + \alpha_2\beta_2 c_{t-1} - \beta_2(1 - \alpha_1)y_{t-1} + \beta_2\varepsilon_{t1} + (1 - \alpha_1)\varepsilon_{t2}, \\ Ay_t &= \alpha_0 + \beta_0 + \beta_1 r_t + g_t + \alpha_2 c_{t-1} - \beta_2 y_{t-1} + \varepsilon_{t1} + \varepsilon_{t2}, \end{aligned}$$

where $A = 1 - \alpha_1 - \beta_2$. Note that the reduced form preserves the equilibrium condition.

The preceding two examples illustrate systems in which there are **behavioral equations** and **equilibrium conditions.** The latter are distinct in that even in an econometric model, they have no disturbances. Another model, which illustrates nearly all the concepts to be discussed in this chapter, is shown in the next example.

Example 15.2 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

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

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 data for 1921 to 1941. The data are listed in Appendix Table F15.1.

15.2.2 ENDOGENEITY AND CAUSALITY

The distinction between "exogenous" and "endogenous" variables in a model is a subtle and sometimes controversial complication. It is the subject of a long literature.[5] We have drawn the distinction in a useful economic fashion at a few points in terms of whether a variable in the model could reasonably be expected to vary "autonomously," independently of the other variables in the model. Thus, in a model of supply and demand, the weather variable in a supply equation seems obviously to be exogenous in a pure sense to the determination of price and quantity, whereas the current price clearly is "endogenous" by any reasonable construction. Unfortunately, this neat classification is of fairly limited use in macroeconomics, where almost no variable can be said to be truly exogenous in the fashion that most observers would understand the term. To take a common example, the estimation of consumption functions by ordinary least squares, as we did in some earlier examples, is usually treated as a respectable enterprise, even though most macroeconomic models (including the examples given here) depart from a consumption function in which income is exogenous. This departure has led analysts, for better or worse, to draw the distinction largely on statistical grounds.

The methodological development in the literature has produced some consensus on this subject. As we shall see, the definitions formalize the economic characterization we drew earlier. We will loosely sketch a few results here for purposes of our derivations to follow. The interested reader is referred to the literature (and forewarned of some challenging reading).

[5]See, for example, Zellner (1979), Sims (1977), Granger (1969), and especially Engle, Hendry, and Richard (1983).

Engle, Hendry, and Richard (1983) define a set of variables $\mathbf{x}_t$ in a parameterized model to be **weakly exogenous** if the full model can be written in terms of a marginal probability distribution for $\mathbf{x}_t$ and a conditional distribution for $\mathbf{y}_t \mid \mathbf{x}_t$ such that estimation of the parameters of the conditional distribution is no less efficient than estimation of the full set of parameters of the joint distribution. This case will be true if none of the parameters in the conditional distribution appears in the marginal distribution for $\mathbf{x}_t$. In the present context, we will need this sort of construction to derive reduced forms the way we did previously.

With reference to time-series applications (although the notion extends to cross sections as well), variables $\mathbf{x}_t$ are said to be **predetermined** in the model if $\mathbf{x}_t$ is independent of all *subsequent* structural disturbances ε_{t+s} for $s > 0$. Variables that are predetermined in a model can be treated, at least asymptotically, as if they were exogenous in the sense that consistent estimates can be obtained when they appear as regressors. We used this result in Chapters 5 and 12 as well, when we derived the properties of regressions containing lagged values of the dependent variable.

A related concept is **Granger causality.** Granger causality (a kind of statistical feedback) is absent when $f(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{y}_{t-1})$ equals $f(\mathbf{x}_t \mid \mathbf{x}_{t-1})$. The definition states that in the conditional distribution, lagged values of $\mathbf{y}_t$ add no information to explanation of movements of $\mathbf{x}_t$ beyond that provided by lagged values of $\mathbf{x}_t$ itself. This concept is useful in the construction of forecasting models. Finally, if $\mathbf{x}_t$ is weakly exogenous and if $\mathbf{y}_{t-1}$ does not Granger cause $\mathbf{x}_t$, then $\mathbf{x}_t$ is **strongly exogenous.**

15.2.3 A GENERAL NOTATION FOR LINEAR SIMULTANEOUS EQUATIONS MODELS[6]

The **structural form** of the model is[7]

$$
\begin{aligned}
\gamma_{11}y_{t1} + \gamma_{21}y_{t2} + \cdots + \gamma_{M1}y_{tM} + \beta_{11}x_{t1} + \cdots + \beta_{K1}x_{tK} &= \varepsilon_{t1},\\
\gamma_{12}y_{t1} + \gamma_{22}y_{t2} + \cdots + \gamma_{M2}y_{tM} + \beta_{12}x_{t1} + \cdots + \beta_{K2}x_{tK} &= \varepsilon_{t2},\\
&\vdots\\
\gamma_{1M}y_{t1} + \gamma_{2M}y_{t2} + \cdots + \gamma_{MM}y_{tM} + \beta_{1M}x_{t1} + \cdots + \beta_{KM}x_{tK} &= \varepsilon_{tM}.
\end{aligned}
\tag{15-2}
$$

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. 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, $t = 1, \ldots, T$.

[6]We will be restricting our attention to linear models in this chapter. **Nonlinear systems** occupy another strand of literature in this area. Nonlinear systems bring forth numerous complications beyond those discussed here and are beyond the scope of this text. Gallant (1987), Gallant and Holly (1980), Gallant and White (1988), Davidson and MacKinnon (1993), and Wooldridge (2002) provide further discussion.

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

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, whereas each row applies to a specific variable.

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. Since 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 **triangular.** In this case, the model is of the form

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

The joint determination of the variables in this model is **recursive.** 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 determining $\mathbf{y}_t$ in terms of $\mathbf{x}_t$ and $\boldsymbol{\varepsilon}_t$ is the **reduced form** of the model,

$$\mathbf{y}_t' = [x_1 \quad x_2 \quad \cdots \quad x_K]_t \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1M} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2M} \\ & & \vdots & \\ \pi_{K1} & \pi_{K2} & \cdots & \pi_{KM} \end{bmatrix} + [\nu_1 \quad \cdots \quad \nu_M]_t$$

$$= -\mathbf{x}_t'\mathbf{B}\boldsymbol{\Gamma}^{-1} + \boldsymbol{\varepsilon}_t'\boldsymbol{\Gamma}^{-1}$$

$$= \mathbf{x}_t'\boldsymbol{\Pi} + \mathbf{v}_t'.$$

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

Example 15.3 Structure and Reduced Form

For the small model in Example 15.1, $\mathbf{y}' = [c, i, y]$, $\mathbf{x}' = [1, r, g, c_{-1}, y_{-1}]$, and

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

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

where $\Delta = 1 - \alpha_1 - \beta_2$. The completeness condition is that α_1 and β_2 do not sum to one.

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

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

For the present, we assume that

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

Later, we will drop this assumption to allow for heteroscedasticity and autocorrelation. 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

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

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

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

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

Under general conditions, we can strengthen this structure to

$$\text{plim}[(1/T)\mathbf{E}'\mathbf{E}] = \boldsymbol{\Sigma}.$$

An important assumption, comparable with the one made in Chapter 5 for the classical regression model, is

$$\text{plim}(1/T)\mathbf{X}'\mathbf{X} = \mathbf{Q}, \quad \text{a finite positive definite matrix.} \tag{15-3}$$

We also assume that

$$\text{plim}(1/T)\mathbf{X}'\mathbf{E} = \mathbf{0}. \tag{15-4}$$

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

$$\text{plim}\,\frac{1}{T}\begin{bmatrix} \mathbf{Y}' \\ \mathbf{X}' \\ \mathbf{V}' \end{bmatrix}[\mathbf{Y} \quad \mathbf{X} \quad \mathbf{V}] = \begin{bmatrix} \boldsymbol{\Pi}'\mathbf{Q}\boldsymbol{\Pi} + \boldsymbol{\Omega} & \boldsymbol{\Pi}'\mathbf{Q} & \boldsymbol{\Omega} \\ \mathbf{Q}\boldsymbol{\Pi} & \mathbf{Q} & \mathbf{0}' \\ \boldsymbol{\Omega} & \mathbf{0} & \boldsymbol{\Omega} \end{bmatrix}. \tag{15-5}$$

15.3 THE PROBLEM OF IDENTIFICATION

Solving the problem to be considered here, the identification problem, logically precedes estimation. We ask at this point whether there is *any* way to obtain estimates of the parameters of the model. We have in hand a certain amount of information upon which to base any inference about its underlying structure. If more than one theory is consistent with the same "data," then the theories are said to be **observationally equivalent** and there is no way of distinguishing them. The structure is said to be *unidentified*.[8]

[8]A useful survey of this issue is Hsiao (1983).

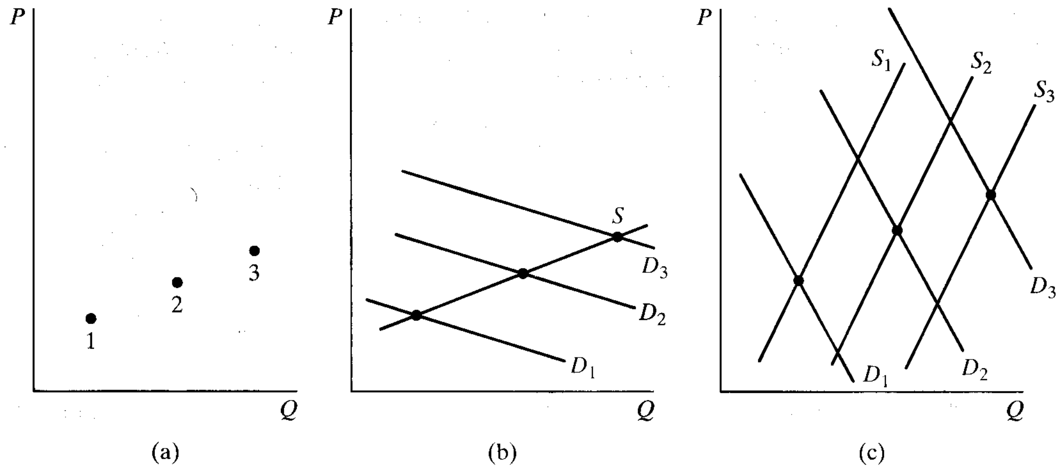

FIGURE 15.1 Market Equilibria.

***Example 15.4 Observational Equivalence*[9]**
The *observed* data consist of the market outcomes shown in Figure 15.1a. We have no knowledge of the conditions of supply and demand beyond our belief that the data represent *equilibria*. Unfortunately, parts (b) and (c) of Figure 15.1 both show *structures*—that is, true underlying supply and demand curves—which are consistent with the data in Figure 15.1a. With only the data in Figure 15.1a, we have no way of determining which of theories 15.1b or c is the right one. Thus, the structure underlying the data in Figure 15.1a is unidentified. To suggest where our discussion is headed, suppose that we add to the preceding the known fact that the conditions of supply were unchanged during the period over which the data were drawn. This rules out 15.1c and identifies 15.1b as the correct structure. Note how this scenario relates to Example 15.1 and to the discussion following that example.

The identification problem is not one of sampling properties or the size of the sample. To focus ideas, it is even useful to suppose that we have at hand an infinite-sized sample of observations on the variables in the model. Now, with this sample and our prior theory, what information do we have? In the reduced form,

$$\mathbf{y}_t' = \mathbf{x}_t'\boldsymbol{\Pi} + \mathbf{v}_t', \qquad E[\mathbf{v}_t\mathbf{v}_t' \,|\, \mathbf{x}_t] = \boldsymbol{\Omega},$$

the predetermined variables are uncorrelated with the disturbances. Thus, we can "observe"

$$\text{plim}(1/T)\mathbf{X}'\mathbf{X} = \mathbf{Q} \text{ [assumed; see (15-3)]},$$
$$\text{plim}(1/T)\mathbf{X}'\mathbf{Y} = \text{plim}(1/T)\mathbf{X}'(\mathbf{X}\boldsymbol{\Pi} + \mathbf{V}) = \mathbf{Q}\boldsymbol{\Pi},$$
$$\text{plim}(1/T)\mathbf{Y}'\mathbf{Y} = \text{plim}(1/T)(\boldsymbol{\Pi}'\mathbf{X}' + \mathbf{V}')(\mathbf{X}\boldsymbol{\Pi} + \mathbf{V}) = \boldsymbol{\Pi}'\mathbf{Q}\boldsymbol{\Pi} + \boldsymbol{\Omega}.$$

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

$$\boldsymbol{\Pi} = \left[\text{plim}\left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)\right]^{-1}\left[\text{plim}\left(\frac{\mathbf{X}'\mathbf{Y}}{T}\right)\right].$$

[9]This example paraphrases the classic argument of Working (1926).

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

$$\mathbf{\Omega} = \text{plim}\,\frac{\mathbf{Y}'\mathbf{Y}}{T} - \text{plim}\left[\frac{\mathbf{Y}'\mathbf{X}}{T}\right]\left[\frac{\mathbf{X}'\mathbf{X}}{T}\right]^{-1}\left[\frac{\mathbf{X}'\mathbf{Y}}{T}\right].$$

This result should be recognized as 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.[10] Now, can we deduce the structural parameters from the reduced form?

The correspondence between the structural and reduced-form parameters is the relationships

$$\mathbf{\Pi} = -\mathbf{B}\mathbf{\Gamma}^{-1} \quad \text{and} \quad \mathbf{\Omega} = E[\mathbf{v}\mathbf{v}'] = (\mathbf{\Gamma}^{-1})'\mathbf{\Sigma}\mathbf{\Gamma}^{-1}.$$

If $\mathbf{\Gamma}$ were known, then we could deduce $\mathbf{B}$ as $-\mathbf{\Pi}\mathbf{\Gamma}$ and $\mathbf{\Sigma}$ as $\mathbf{\Gamma}'\mathbf{\Omega}\mathbf{\Gamma}$. It would appear, therefore, that our problem boils down to obtaining $\mathbf{\Gamma}$, which makes sense. If $\mathbf{\Gamma}$ were known, then we could rewrite (15-2), collecting the endogenous variables times their respective coefficients on the left-hand side of a regression, and estimate the remaining unknown coefficients on the predetermined variables by ordinary least squares.[11]

The identification question we will pursue can be posed as follows: We can "observe" the reduced form. We must deduce the structure from what we know about the reduced form. If there is more than one structure that can lead to the same reduced form, then we cannot say that we can "estimate the structure." Which structure would that be? Suppose that the "true" structure is $[\mathbf{\Gamma}, \mathbf{B}, \mathbf{\Sigma}]$. Now consider a different structure, $\mathbf{y}'\tilde{\mathbf{\Gamma}} + \mathbf{x}'\tilde{\mathbf{B}} = \tilde{\boldsymbol{\varepsilon}}'$, that is obtained by postmultiplying the first structure by some nonsingular matrix $\mathbf{F}$. Thus, $\tilde{\mathbf{\Gamma}} = \mathbf{\Gamma}\mathbf{F}$, $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{F}$, $\tilde{\boldsymbol{\varepsilon}}' = \boldsymbol{\varepsilon}'\mathbf{F}$. The reduced form that corresponds to this new structure is, unfortunately, the same as the one that corresponds to the old one;

$$\tilde{\mathbf{\Pi}} = -\tilde{\mathbf{B}}\tilde{\mathbf{\Gamma}}^{-1} = -\mathbf{B}\mathbf{F}\mathbf{F}^{-1}\mathbf{\Gamma}^{-1} = \mathbf{\Pi},$$

and, in the same fashion, $\tilde{\mathbf{\Omega}} = \mathbf{\Omega}$. The false structure looks just like the true one, at least in terms of the information we have. Statistically, there is no way we can tell them apart. The structures are observationally equivalent.

Since $\mathbf{F}$ was chosen arbitrarily, we conclude that *any* nonsingular transformation of the original structure has the same reduced form. Any reason for optimism that we might have had should be abandoned. As the model stands, there is no means by which the structural parameters can be deduced from the reduced form. The practical implication is that if the only information that we have is the reduced-form parameters, then the structural model is not estimable. So how were we able to identify the models

[10] We have not necessarily shown that this is *all* the information in the sample. In general, we observe the conditional distribution $f(\mathbf{y}_t \mid \mathbf{x}_t)$, which constitutes the likelihood for the reduced form. With normally distributed disturbances, this distribution is a function of $\mathbf{\Pi}$, $\mathbf{\Omega}$. (See Section 15.6.2.) With other distributions, other or higher moments of the variables might provide additional information. See, for example, Goldberger (1964, p. 311), Hausman (1983, pp. 402–403), and especially Riersøl (1950).

[11] This method is precisely the approach of the LIML estimator. See Section 15.5.5.

in the earlier examples? The answer is by bringing to bear our **nonsample information,** namely our theoretical restrictions. Consider the following examples:

Example 15.5 Identification

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,

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

The reduced form is

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

With only four reduced-form coefficients and six structural parameters, it is obvious that there will not be a complete solution for all six structural parameters in terms of the four reduced 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 restriction identifies the supply parameters. But this step is as far as we can go.

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

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

The structure is now

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

The reduced form is

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

where $\boldsymbol{\Delta} = (\alpha_1 - \beta_1)$. Every false structure has the same reduced form. But in the coefficient matrix,

$$\tilde{\mathbf{B}} = \mathbf{BF} = \begin{bmatrix} \alpha_0 f_{11} + \beta_0 f_{12} & \alpha_0 f_{12} + \beta_0 f_{22} \\ \alpha_2 f_{11} & \alpha_2 f_{12} \\ \beta_2 f_{21} & \beta_2 f_{22} \end{bmatrix},$$

if f_{12} is not zero, then the imposter will have income appearing in the supply equation, which our theory has ruled out. Likewise, if f_{21} is not zero, then z will appear in the demand equation, which is also ruled out by our theory. Thus, although all false structures have the same reduced form as the true one, the only one that is consistent with our theory (i.e., is **admissible**) and has coefficients of 1 on q in both equations (examine $\mathbf{\Gamma F}$) is $\mathbf{F} = \mathbf{I}$. This transformation just produces the original structure.

The unique solutions for the structural parameters in terms of the reduced-form parameters are

$$\alpha_0 = \pi_{11} - \pi_{12}\left(\frac{\pi_{31}}{\pi_{32}}\right), \quad \beta_0 = \pi_{11} - \pi_{12}\left(\frac{\pi_{21}}{\pi_{22}}\right),$$

$$\alpha_1 = \frac{\pi_{31}}{\pi_{32}}, \quad \beta_1 = \frac{\pi_{21}}{\pi_{22}},$$

$$\alpha_2 = \pi_{22}\left(\frac{\pi_{21}}{\pi_{22}} - \frac{\pi_{31}}{\pi_{32}}\right), \quad \beta_2 = \pi_{32}\left(\frac{\pi_{31}}{\pi_{32}} - \frac{\pi_{21}}{\pi_{22}}\right).$$

The preceding discussion has considered two equivalent methods of establishing identifiability. If it is possible to deduce the structural parameters from the known reduced form parameters, then the model is identified. Alternatively, if it can be shown that no false structure is admissible—that is, satisfies the theoretical restrictions—then the model is identified.[12]

15.3.1 THE RANK AND ORDER CONDITIONS FOR IDENTIFICATION

It is useful to summarize what we have determined thus far. The unknown structural parameters consist of

$$\mathbf{\Gamma} = \text{an } M \times M \text{ nonsingular matrix},$$

$$\mathbf{B} = \text{a } K \times M \text{ parameter matrix},$$

$$\mathbf{\Sigma} = \text{an } M \times M \text{ symmetric positive definite matrix}.$$

The known, reduced-form parameters are

$$\mathbf{\Pi} = \text{a } K \times M \text{ reduced-form coefficients matrix},$$

$$\mathbf{\Omega} = \text{an } M \times M \text{ reduced-form covariance matrix}.$$

Simply counting parameters in the structure and reduced forms yields an excess of

$$l = M^2 + KM + \tfrac{1}{2}M(M+1) - KM - \tfrac{1}{2}M(M+1) = M^2,$$

which is, as might be expected from the earlier results, the number of unknown elements in $\mathbf{\Gamma}$. Without further information, identification is clearly impossible. The additional information comes in several forms.

1. *Normalizations.* In each equation, one variable has a coefficient of 1. This normalization is a necessary scaling of the equation that is logically equivalent to putting one variable on the left-hand side of a regression. For purposes of identification (and some estimation methods), the choice among the endogenous variables is arbitrary. But at the time the model is formulated, each equation will usually have some natural dependent variable. The normalization does not identify the dependent variable in any formal or causal sense. For example, in a model of supply and demand, both the "demand"

[12] For other interpretations, see Amemiya (1985, p. 230) and Gabrielsen (1978). Some deeper theoretical results on identification of parameters in econometric models are given by Bekker and Wansbeek (2001).

equation, $Q = f(P, \mathbf{x})$, and the "inverse demand" equation, $P = g(Q, \mathbf{x})$, are appropriate specifications of the relationship between price and quantity. We note, though, the following:

> With the normalizations, there are $M(M-1)$, not M^2, undetermined values in $\mathbf{\Gamma}$ and this many indeterminacies in the model to be resolved through nonsample information.

2. *Identities.* In some models, variable definitions or equilibrium conditions imply that all the coefficients in a particular equation are known. In the preceding market example, there are three equations, but the third is the equilibrium condition $Q_d = Q_s$. Klein's Model I (Example 15.3) contains six equations, including two accounting identities and the equilibrium condition. There is no question of identification with respect to identities. They may be carried as additional equations in the model, as we do with Klein's Model I in several later examples, or built into the model a priori, as is typical in models of supply and demand.

The substantive nonsample information that will be used in identifying the model will consist of the following:

3. *Exclusions.* The omission of variables from an equation places zeros in $\mathbf{B}$ and $\mathbf{\Gamma}$. In Example 15.5, the exclusion of income from the supply equation served to identify its parameters.

4. *Linear restrictions.* Restrictions on the structural parameters may also serve to rule out false structures. For example, a long-standing problem in the estimation of production models using time-series data is the inability to disentangle the effects of economies of scale from those of technological change. In some treatments, the solution is to assume that there are constant returns to scale, thereby identifying the effects due to technological change.

5. *Restrictions on the disturbance covariance matrix.* In the identification of a model, these are similar to restrictions on the slope parameters. For example, if the previous market model were to apply to a microeconomic setting, then it would probably be reasonable to assume that the structural disturbances in these supply and demand equations are uncorrelated. Section 15.3.3 shows a case in which a covariance restriction identifies an otherwise unidentified model.

To formalize the identification criteria, we require a notation for a single equation. The coefficients of the jth equation are contained in the jth columns of $\mathbf{\Gamma}$ and $\mathbf{B}$. The jth equation is

$$\mathbf{y}'\mathbf{\Gamma}_j + \mathbf{x}'\mathbf{B}_j = \boldsymbol{\varepsilon}_j. \qquad \textbf{(15-6)}$$

(For convenience, we have dropped the observation subscript.) In this equation, we know that (1) one of the elements in $\mathbf{\Gamma}_j$ is one and (2) some variables that appear elsewhere in the model are excluded from this equation. Table 15.1 defines the notation used to incorporate these restrictions in (15-6).

Equation j may be written

$$\mathbf{y}_j = \mathbf{Y}_j'\boldsymbol{\gamma}_j + \mathbf{Y}_j^{*\prime}\boldsymbol{\gamma}_j^* + \mathbf{x}_j'\boldsymbol{\beta}_j + \mathbf{x}_j^{*\prime}\boldsymbol{\beta}_j^* + \boldsymbol{\varepsilon}_j.$$

TABLE 15.1 Components of Equation j (Dependent Variable $= y_j$)

	Endogenous Variables	*Exogenous Variables*
Included	$\mathbf{Y}_j = M_j$ variables	$\mathbf{x}_j = K_j$ variables
Excluded	$\mathbf{Y}_j^* = M_j^*$ variables	$\mathbf{x}_j^* = K_j^*$ variables

The number of equations is $M_j + M_j^* + 1 = M$.
The number of exogenous variables is $K_j + K_j^* = K$.
The coefficient on y_j in equation j is 1.
$*$s will always be associated with excluded variables.

The exclusions imply that $\boldsymbol{\gamma}_j^* = \mathbf{0}$ and $\boldsymbol{\beta}_j^* = \mathbf{0}$. Thus,

$$\boldsymbol{\Gamma}_j' = [1 \;\; -\boldsymbol{\gamma}_j' \;\; \mathbf{0}'] \quad \text{and} \quad \mathbf{B}_j' = [-\boldsymbol{\beta}_j' \;\; \mathbf{0}'].$$

(Note the sign convention.) For this equation, we partition the reduced-form coefficient matrix in the same fashion:

$$[\mathbf{y}_j \;\; \mathbf{Y}_j' \;\; \mathbf{Y}_j^{*\prime}] = [\mathbf{x}_j' \;\; \mathbf{x}_j^{*\prime}] \overset{\quad(1)\quad (M_j)\quad (M_j^*)}{\begin{bmatrix} \boldsymbol{\pi}_j & \underline{\boldsymbol{\Pi}}_j & \bar{\boldsymbol{\Pi}}_j \\ \boldsymbol{\pi}_j^* & \boldsymbol{\Pi}_j^* & \bar{\boldsymbol{\Pi}}_j^* \end{bmatrix}} + [\mathbf{v}_j \;\; \mathbf{V}_j \;\; \mathbf{V}_j^*] \quad \begin{matrix} [K_j \text{ rows}] \\ [K_j^* \text{ rows}]. \end{matrix} \tag{15-7}$$

The reduced-form coefficient matrix is

$$\boldsymbol{\Pi} = -\mathbf{B}\boldsymbol{\Gamma}^{-1},$$

which implies that

$$\boldsymbol{\Pi}\boldsymbol{\Gamma} = -\mathbf{B}.$$

The jth column of this matrix equation applies to the jth equation,

$$\boldsymbol{\Pi}\boldsymbol{\Gamma}_j = -\mathbf{B}_j.$$

Inserting the parts from Table 15.1 yields

$$\begin{bmatrix} \boldsymbol{\pi}_j & \underline{\boldsymbol{\Pi}}_j & \bar{\boldsymbol{\Pi}}_j \\ \boldsymbol{\pi}_j^* & \boldsymbol{\Pi}_j^* & \bar{\boldsymbol{\Pi}}_j^* \end{bmatrix} \begin{bmatrix} 1 \\ -\boldsymbol{\gamma}_j \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_j \\ \mathbf{0} \end{bmatrix}.$$

Now extract the two subequations,

$$\boldsymbol{\pi}_j - \underline{\boldsymbol{\Pi}}_j \boldsymbol{\gamma}_j = \boldsymbol{\beta}_j \quad (K_j \text{ equations}), \tag{15-8}$$

$$\boldsymbol{\pi}_j^* - \boldsymbol{\Pi}_j^* \boldsymbol{\gamma}_j = \mathbf{0} \quad (K_j^* \text{ equations}), \tag{15-9}$$

$$(1) \quad (M_j).$$

The solution for $\mathbf{B}$ in terms of $\boldsymbol{\Gamma}$ that we observed at the beginning of this discussion is in (15-8). Equation (15-9) may be written

$$\boldsymbol{\Pi}_j^* \boldsymbol{\gamma}_j = \boldsymbol{\pi}_j^*. \tag{15-10}$$

This system is K_j^* equations in M_j unknowns. If they can be solved for $\boldsymbol{\gamma}_j$, then (15-8) gives the solution for $\boldsymbol{\beta}_j$ and the equation is identified. For there to be a solution, there must be at least as many equations as unknowns, which leads to the following condition.

DEFINITION 15.1 Order Condition for Identification of Equation *j*

$$K_j^* \geq M_j. \tag{15-11}$$

The number of exogenous variables excluded from equation j must be at least as large as the number of endogenous variables included in equation j.

The order condition is only a counting rule. It is a necessary but not sufficient condition for identification. It ensures that (15-10) has at least one solution, but it does not ensure that it has only one solution. The sufficient condition for uniqueness follows.

DEFINITION 15.2 Rank Condition for Identification

$$\text{rank}[\pi_j^*, \Pi_j^*] = \text{rank}[\Pi_j^*] = M_j.$$

This condition imposes a restriction on a submatrix of the reduced-form coefficient matrix.

The rank condition ensures that there is exactly one solution for the structural parameters given the reduced-form parameters. Our alternative approach to the identification problem was to use the prior restrictions on $[\Gamma, \mathbf{B}]$ to eliminate all false structures. An equivalent condition based on this approach is simpler to apply and has more intuitive appeal. We first rearrange the structural coefficients in the matrix

$$\mathbf{A} = \begin{bmatrix} \Gamma \\ \mathbf{B} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{A}_1 \\ -\gamma_j & \mathbf{A}_2 \\ \mathbf{0} & \mathbf{A}_3 \\ -\beta_j & \mathbf{A}_4 \\ \mathbf{0} & \mathbf{A}_5 \end{bmatrix} = [\mathbf{a}_j \quad \mathbf{A}_j]. \tag{15-12}$$

The jth column in a false structure $[\Gamma\mathbf{F}, \mathbf{BF}]$ (i.e., the imposter for our equation j) would be $[\Gamma\mathbf{f}_j, \mathbf{Bf}_j]$, where $\mathbf{f}_j$ is the jth column of $\mathbf{F}$. This new jth equation is to be built up as a linear combination of the old one and the other equations in the model. Thus, partitioning as previously,

$$\tilde{\mathbf{a}}_j = \begin{bmatrix} 1 & \mathbf{A}_1 \\ -\gamma_j & \mathbf{A}_2 \\ \mathbf{0} & \mathbf{A}_3 \\ -\beta_j & \mathbf{A}_4 \\ \mathbf{0} & \mathbf{A}_5 \end{bmatrix} \begin{bmatrix} f^0 \\ \mathbf{f}^1 \end{bmatrix} = \begin{bmatrix} 1 \\ \tilde{\gamma}_j \\ \mathbf{0} \\ \tilde{\beta}_j \\ \mathbf{0} \end{bmatrix}.$$

If this hybrid is to have the same variables as the original, then it must have nonzero elements in the same places, which can be ensured by taking $f^0 = 1$, and zeros in the same positions as the original $\mathbf{a}_j$. Extracting the third and fifth blocks of rows, if $\tilde{\mathbf{a}}_j$ is to be admissible, then it must meet the requirement

$$\begin{bmatrix} \mathbf{A}_3 \\ \mathbf{A}_5 \end{bmatrix} \mathbf{f}^1 = \mathbf{0}.$$

This equality is not possible if the $(M_j^* + K_j^*) \times (M-1)$ matrix in brackets has full column rank, so we have the equivalent rank condition,

$$\operatorname{rank} \begin{bmatrix} \mathbf{A}_3 \\ \mathbf{A}_5 \end{bmatrix} = M - 1.$$

The corresponding order condition is that the matrix in brackets must have at least as many rows as columns. Thus, $M_j^* + K_j^* \geq M - 1$. But since $M = M_j + M_j^* + 1$, this condition is the same as the order condition in (15-11). The equivalence of the two rank conditions is pursued in the exercises.

The preceding provides a simple method for checking the rank and order conditions. We need only arrange the structural parameters in a tableau and examine the relevant submatrices one at a time; $\mathbf{A}_3$ and $\mathbf{A}_5$ are the structural coefficients in the other equations on the variables that are excluded from equation j.

One rule of thumb is sometimes useful in checking the rank and order conditions of a model: *If every equation has its own predetermined variable, the entire model is identified.* The proof is simple and is left as an exercise. For a final example, we consider a somewhat larger model.

Example 15.6 Identification of Klein's Model I

The structural coefficients in the six equations of Klein's Model I, transposed and multiplied by −1 for convenience, are listed in Table 15.2. Identification of the consumption function requires that the matrix $[\mathbf{A}_3', \mathbf{A}_5']$ have rank 5. The columns of this matrix are contained in boxes in the table. None of the columns indicated by arrows can be formed as linear combinations of the other columns, so the rank condition is satisfied. Verification of the rank and order conditions for the other two equations is left as an exercise.

It is unusual for a model to pass the order but not the rank condition. Generally, either the conditions are obvious or the model is so large and has so many predetermined

TABLE 15.2 Klein's Model I, Structural Coefficients

	Γ'						B'							
	C	I	W^p	X	P	K	1	W^g	G	T	A	P_{-1}	K_{-1}	X_{-1}
C	−1	0	α_3	0	α_1	0	α_0	α_3	0	0	0	α_2	0	0
I	0	−1	0	0	β_1	0	β_0	0	0	0	0	β_2	β_3	0
W^p	0	0	−1	γ_1	0	0	γ_0	0	0	0	γ_3	0	0	γ_2
X	1	1	0	−1	0	0	0	0	1	0	0	0	0	0
P	0	0	−1	1	−1	0	0	0	0	−1	0	0	0	0
K	0	1	0	0	0	−1	0	0	0	0	0	0	1	0
				↑		↑				↑	↑		↑	
			$\mathbf{A}_3'$						$\mathbf{A}_5'$					

variables that the conditions are met trivially. In practice, it is simple to check both conditions for a small model. For a large model, frequently only the order condition is verified. We distinguish three cases:

1. *Underidentified.* $K_j^* < M_j$ or rank condition fails.
2. *Exactly identified.* $K_j^* = M_j$ and rank condition is met.
3. *Overidentified.* $K_j^* > M_j$ and rank condition is met.

15.3.2 IDENTIFICATION THROUGH OTHER NONSAMPLE INFORMATION

The rank and order conditions given in the preceding section apply to identification of an equation through **exclusion restrictions.** Intuition might suggest that other types of nonsample information should be equally useful in securing identification. To take a specific example, suppose that in Example 15.5, it is known that β_2 equals 2, not 0. The second equation could then be written as

$$\mathbf{q}_s - 2\mathbf{z} = \mathbf{q}_s^* = \beta_0 + \beta_1\mathbf{p} + \beta_j^*\mathbf{z} + \varepsilon_2.$$

But we know that $\beta_j^* = 0$, so the supply equation is identified by this restriction. As this example suggests, a linear restriction on the parameters *within* an equation is, for identification purposes, essentially the same as an exclusion.[13] By an appropriate manipulation—that is, by "solving out" the restriction—we can turn the restriction into one more exclusion. The order condition that emerges is

$$n_j \geq M - 1,$$

where n_j is the total number of restrictions. Since $M - 1 = M_j + M_j^*$ and n_j is the number of exclusions plus r_j, the number of additional restrictions, this condition is equivalent to

$$r_j + K_j^* + M_j^* \geq M_j + M_j^*$$

or

$$r_j + K_j^* \geq M_j.$$

This result is the same as (15-11) save for the addition of the number of restrictions, which is the result suggested previously.

15.3.3 IDENTIFICATION THROUGH COVARIANCE RESTRICTIONS—THE FULLY RECURSIVE MODEL

The observant reader will have noticed that no mention of $\boldsymbol{\Sigma}$ is made in the preceding discussion. To this point, all the information provided by $\boldsymbol{\Omega}$ is used in the estimation of $\boldsymbol{\Sigma}$; for given $\boldsymbol{\Gamma}$, the relationship between $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ is one-to-one. Recall that $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}'\boldsymbol{\Omega}\boldsymbol{\Gamma}$. But if restrictions are placed on $\boldsymbol{\Sigma}$, then there is more information in $\boldsymbol{\Omega}$ than is needed for estimation of $\boldsymbol{\Sigma}$. The excess information can be used instead to help infer the elements

[13]The analysis is more complicated if the restrictions are *across* equations, that is, involve the parameters in more than one equation. Kelly (1975) contains a number of results and examples.

in Γ. A useful case is that of zero covariances across the disturbances.[14] Once again, it is most convenient to consider this case in terms of a false structure. If the structure is $[\mathbf{\Gamma}, \mathbf{B}, \mathbf{\Sigma}]$, then a false structure would have parameters

$$[\tilde{\mathbf{\Gamma}}, \tilde{\mathbf{B}}, \tilde{\mathbf{\Sigma}}] = [\mathbf{\Gamma F}, \mathbf{BF}, \mathbf{F}'\mathbf{\Sigma F}].$$

If any of the elements in $\mathbf{\Sigma}$ are zero, then the false structure must preserve those restrictions to be admissible. For example, suppose that we specify that $\sigma_{12} = 0$. Then it must also be true that $\tilde{\sigma}_{12} = \mathbf{f}_1'\mathbf{\Sigma f}_2 = 0$, where $\mathbf{f}_1$ and $\mathbf{f}_2$ are columns of $\mathbf{F}$. As such, there is a restriction on $\mathbf{F}$ that may identify the model.

The **fully recursive model** is an important special case of the preceding result. A **triangular system** is

$$\begin{aligned}
y_1 &= \boldsymbol{\beta}_1'\mathbf{x} + \varepsilon_1, \\
y_2 &= \gamma_{12}y_1 + \boldsymbol{\beta}_2'\mathbf{x} + \varepsilon_2, \\
&\vdots \\
y_M &= \gamma_{1M}y_1 + \gamma_{2M}y_2 + \cdots + \gamma_{M-1,M}y_{M-1} + \boldsymbol{\beta}_M'\mathbf{x} + \varepsilon_M.
\end{aligned}$$

We place no restrictions on $\mathbf{B}$. The first equation is identified, since it is already in reduced form. But for any of the others, linear combinations of it and the ones above it involve the same variables. Thus, we conclude that *without some identifying restrictions, only the parameters of the first equation in a triangular system are identified*. But suppose that $\mathbf{\Sigma}$ is diagonal. Then the entire model is identified, as we now prove. As usual, we attempt to find a false structure that satisfies the restrictions of the model.

The jth column of $\mathbf{F}$, $\mathbf{f}_j$, is the coefficients in a linear combination of the equations that will be an imposter for equation j. Many $\mathbf{f}_j$'s are already precluded.

1. $\mathbf{f}_1$ must be the first column of an identity matrix. The first equation is identified and normalized on y_1.
2. In all remaining columns of $\mathbf{F}$, all elements below the diagonal must be zero, since an equation can only involve the ys in it or in the equations above it.

Without further restrictions, any upper triangular $\mathbf{F}$ is an admissible transformation. But with a diagonal $\mathbf{\Sigma}$, we have more information. Consider the second column. Since $\tilde{\mathbf{\Sigma}}$ must be diagonal, $\mathbf{f}_1'\mathbf{\Sigma f}_2 = 0$. But given $\mathbf{f}_1$ in 1 above,

$$\mathbf{f}_1'\mathbf{\Sigma f}_2 = \sigma_{11}f_{12} = 0,$$

so $f_{12} = 0$. The second column of $\mathbf{F}$ is now complete and is equal to the second column of $\mathbf{I}$. Continuing in the same manner, we find that

$$\mathbf{f}_1'\mathbf{\Sigma f}_3 = 0 \quad \text{and} \quad \mathbf{f}_2'\mathbf{\Sigma f}_3 = 0$$

will suffice to establish that $\mathbf{f}_3$ is the third column of $\mathbf{I}$. In this fashion, it can be shown that the only admissible $\mathbf{F}$ is $\mathbf{F} = \mathbf{I}$, which was to be shown. With $\mathbf{\Gamma}$ upper triangular, $M(M-1)/2$ unknown parameters remained. That is exactly the number of restrictions placed on $\mathbf{\Sigma}$ when it was assumed to be diagonal.

[14]More general cases are discussed in Hausman (1983) and Judge et al. (1985).

15.4 METHODS OF ESTIMATION

It is possible to estimate the reduced-form parameters, $\mathbf{\Pi}$ and $\mathbf{\Omega}$, consistently by ordinary least squares. But except for forecasting $\mathbf{y}$ given $\mathbf{x}$, these are generally not the parameters of interest; $\mathbf{\Gamma}$, $\mathbf{B}$, and $\mathbf{\Sigma}$ are. The ordinary least squares (OLS) estimators of the structural parameters are inconsistent, ostensibly because the included endogenous variables in each equation are correlated with the disturbances. Still, it is at least of passing interest to examine what is estimated by ordinary least squares, particularly in view of its widespread use (despite its inconsistency). Since the proof of identification was based on solving for $\mathbf{\Gamma}$, $\mathbf{B}$, and $\mathbf{\Sigma}$ from $\mathbf{\Pi}$ and $\mathbf{\Omega}$, one way to proceed is to apply our finding to the sample estimates, $\mathbf{P}$ and $\mathbf{W}$. This **indirect least squares** approach is feasible but inefficient. Worse, there will usually be more than one possible estimator and no obvious means of choosing among them. There are two approaches for direct estimation, both based on the principle of instrumental variables. It is possible to estimate each equation separately using a **limited information** estimator. But the same principle that suggests that joint estimation brings efficiency gains in the seemingly unrelated regressions setting of the previous chapter is at work here, so we shall also consider **full information** or system methods of estimation.

15.5 SINGLE EQUATION: LIMITED INFORMATION ESTIMATION METHODS

Estimation of the system one equation at a time has the benefit of computational simplicity. But because these methods neglect information contained in the other equations, they are labeled limited information methods.

15.5.1 ORDINARY LEAST SQUARES

For all T observations, the nonzero terms in the jth equation are

$$\begin{aligned}\mathbf{y}_j &= \mathbf{Y}_j\boldsymbol{\gamma}_j + \mathbf{X}_j\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j \\ &= \mathbf{Z}_j\boldsymbol{\delta}_j + \boldsymbol{\varepsilon}_j.\end{aligned}$$

The M reduced-form equations are $\mathbf{Y} = \mathbf{X\Pi} + \mathbf{V}$. For the included endogenous variables $\mathbf{Y}_j$, the reduced forms are the M_j appropriate columns of $\mathbf{\Pi}$ and $\mathbf{V}$, written

$$\mathbf{Y}_j = \mathbf{X\Pi}_j + \mathbf{V}_j. \tag{15-13}$$

[Note that $\mathbf{\Pi}_j$ is the middle part of $\mathbf{\Pi}$ shown in (15-7).] Likewise, $\mathbf{V}_j$ is M_j columns of $\mathbf{V} = \mathbf{E\Gamma}^{-1}$. This least squares estimator is

$$\mathbf{d}_j = [\mathbf{Z}_j'\mathbf{Z}_j]^{-1}\mathbf{Z}_j'\mathbf{y}_j = \boldsymbol{\delta}_j + \begin{bmatrix}\mathbf{Y}_j'\mathbf{Y}_j & \mathbf{Y}_j'\mathbf{X}_j \\ \mathbf{X}_j'\mathbf{Y}_j & \mathbf{X}_j'\mathbf{X}_j\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{Y}_j'\boldsymbol{\varepsilon}_j \\ \mathbf{X}_j'\boldsymbol{\varepsilon}_j\end{bmatrix}.$$

None of the terms in the inverse matrix converge to $\mathbf{0}$. Although $\text{plim}(1/T)\mathbf{X}_j'\boldsymbol{\varepsilon}_j = \mathbf{0}$, $\text{plim}(1/T)\mathbf{Y}_j'\boldsymbol{\varepsilon}_j$ is nonzero, which means that both parts of $\mathbf{d}_j$ are inconsistent. (This is the "**simultaneous equations bias**" of least squares.) Although we can say with certainty that $\mathbf{d}_j$ is inconsistent, we cannot state how serious this problem is. OLS does have the virtue of computational simplicity, although with modern software, this virtue is extremely modest. For better or worse, OLS is a very commonly used estimator in this context. We will return to this issue later in a comparison of several estimators.

An intuitively appealing form of simultaneous equations model is the **triangular system** that we examined in Section 15.5.3,

$$
\begin{aligned}
&(1)\ y_1 = \mathbf{x}'\boldsymbol{\beta}_1 && + \varepsilon_1,\\
&(2)\ y_2 = \mathbf{x}'\boldsymbol{\beta}_2 + \gamma_{12} y_1 && + \varepsilon_2,\\
&(3)\ y_3 = \mathbf{x}'\boldsymbol{\beta}_3 + \gamma_{13} y_1 + \gamma_{23} y_2 && + \varepsilon_3,
\end{aligned}
$$

and so on. If $\boldsymbol{\Gamma}$ is triangular and $\boldsymbol{\Sigma}$ is diagonal, so that the disturbances are uncorrelated, then the system is a **fully recursive model.** (No restrictions are placed on $\mathbf{B}$.) It is easy to see that in this case, the entire system may be estimated consistently (and, as we shall show later, efficiently) by ordinary least squares. The first equation is a classical regression model. In the second equation, $\text{Cov}(y_1, \varepsilon_2) = \text{Cov}(\mathbf{x}'\boldsymbol{\beta}_1 + \varepsilon_1, \varepsilon_2) = 0$, so it too may be estimated by ordinary least squares. Proceeding in the same fashion to (3), it is clear that y_1 and ε_3 are uncorrelated. Likewise, if we substitute (1) in (2) and then the result for y_2 in (3), then we find that y_2 is also uncorrelated with ε_3. Continuing in this way, we find that in every equation the full set of right-hand variables is uncorrelated with the respective disturbance. The result is that *the fully recursive model may be consistently estimated using equation-by-equation ordinary least squares.* (In the more general case, in which $\boldsymbol{\Sigma}$ is not diagonal, the preceding argument does not apply.)

15.5.2 ESTIMATION BY INSTRUMENTAL VARIABLES

In the next several sections, we will discuss various methods of consistent and efficient estimation. As will be evident quite soon, there is a surprisingly long menu of choices. It is a useful result that all of the methods in general use can be placed under the umbrella of **instrumental variable (IV) estimators.**

Returning to the structural form, we first consider direct estimation of the jth equation,

$$
\begin{aligned}
\mathbf{y}_j &= \mathbf{Y}_j\boldsymbol{\gamma}_j + \mathbf{X}_j\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j\\
&= \mathbf{Z}_j\boldsymbol{\delta}_j + \boldsymbol{\varepsilon}_j.
\end{aligned}
\tag{15-14}
$$

As we saw previously, the OLS estimator of $\boldsymbol{\delta}_j$ is inconsistent because of the correlation of $\mathbf{Z}_j$ and $\boldsymbol{\varepsilon}_j$. A general method of obtaining **consistent estimates** is the method of instrumental variables. (See Section 5.4.) Let $\mathbf{W}_j$ be a $T \times (M_j + K_j)$ matrix that satisfies the requirements for an IV estimator,

$$\text{plim}(1/T)\mathbf{W}_j'\mathbf{Z}_j = \boldsymbol{\Sigma}_{wz} = \text{a finite nonsingular matrix,} \tag{15-15a}$$

$$\text{plim}(1/T)\mathbf{W}_j'\boldsymbol{\varepsilon}_j = \mathbf{0}, \tag{15-15b}$$

$$\text{plim}(1/T)\mathbf{W}_j'\mathbf{W}_j = \boldsymbol{\Sigma}_{ww} = \text{a positive definite matrix.} \tag{15-15c}$$

Then the IV estimator,

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

will be consistent and have asymptotic covariance matrix

$$\text{Asy. Var}[\hat{\boldsymbol\delta}_{j,\text{IV}}] = \frac{\sigma_{jj}}{T}\text{plim}\left[\frac{1}{T}\mathbf{W}_j'\mathbf{Z}_j\right]^{-1}\left[\frac{1}{T}\mathbf{W}_j'\mathbf{W}_j\right]\left[\frac{1}{T}\mathbf{Z}_j'\mathbf{W}_j\right]^{-1}$$
$$= \frac{\sigma_{jj}}{T}\left[\boldsymbol\Sigma_{wz}^{-1}\boldsymbol\Sigma_{ww}\boldsymbol\Sigma_{zw}^{-1}\right]. \qquad \textbf{(15-16)}$$

A consistent estimator of σ_{jj} is

$$\hat{\sigma}_{jj} = \frac{(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol\delta}_{j,\text{IV}})'(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol\delta}_{j,\text{IV}})}{T}, \qquad \textbf{(15-17)}$$

which is the familiar sum of squares of the estimated disturbances. A degrees of freedom correction for the denominator, $T - M_j - K_j$, is sometimes suggested. Asymptotically, the correction is immaterial. Whether it is beneficial in a small sample remains to be settled. The resulting estimator is not unbiased in any event, as it would be in the classical regression model. In the interest of simplicity (only), we shall omit the degrees of freedom correction in what follows. Current practice in most applications is to make the correction.

The various estimators that have been developed for simultaneous-equations models are all IV estimators. They differ in the choice of instruments and in whether the equations are estimated one at a time or jointly. We divide them into two classes, **limited information** or **full information,** on this basis.

15.5.3 TWO-STAGE LEAST SQUARES

The method of two-stage least squares is the most common method used for estimating simultaneous-equations models. We developed the full set of results for this estimator in Section 5.4. By merely changing notation slightly, the results of Section 5.4 are exactly the derivation of the estimator we will describe here. Thus, you might want to review this section before continuing.

The **two-stage least squares (2SLS)** method consists of using as the instruments for $\mathbf{Y}_j$ the predicted values in a regression of $\mathbf{Y}_j$ on *all* the xs in the system:

$$\hat{\mathbf{Y}}_j = \mathbf{X}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_j] = \mathbf{X}\mathbf{P}_j. \qquad \textbf{(15-18)}$$

It can be shown that absent heteroscedasticity or autocorrelation, this produces the most efficient IV estimator that can be formed using only the columns of $\mathbf{X}$. Note the emulation of $E[\mathbf{Y}_j] = \mathbf{X}\boldsymbol\Pi_j$ in the result. The 2SLS estimator is, thus,

$$\hat{\boldsymbol\delta}_{j,\text{2SLS}} = \begin{bmatrix}\hat{\mathbf{Y}}_j'\mathbf{Y}_j & \hat{\mathbf{Y}}_j'\mathbf{X}_j \\ \mathbf{X}_j'\mathbf{Y}_j & \mathbf{X}_j'\mathbf{X}_j\end{bmatrix}^{-1}\begin{bmatrix}\hat{\mathbf{Y}}_j'\mathbf{y}_j \\ \mathbf{X}_j'\mathbf{y}_j\end{bmatrix}. \qquad \textbf{(15-19)}$$

Before proceeding, it is important to emphasize the role of the identification condition in this result. In the matrix $[\hat{\mathbf{Y}}_j, \mathbf{X}_j]$, which has $M_j + K_j$ columns, all columns are linear functions of the K columns of $\mathbf{X}$. There exist, at most, K linearly independent combinations of the columns of $\mathbf{X}$. If the equation is not identified, then $M_j + K_j$ is greater than K, and $[\hat{\mathbf{Y}}_j, \mathbf{X}_j]$ will not have full column rank. In this case, the 2SLS estimator cannot be computed. If, however, the order condition but not the rank condition is met, then although the 2SLS estimator can be computed, it is not a consistent estimator. There are a few useful simplifications. First, since $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = (\mathbf{I} - \mathbf{M})$ is idempotent, $\hat{\mathbf{Y}}_j'\mathbf{Y}_j = \hat{\mathbf{Y}}_j'\hat{\mathbf{Y}}_j$. Second, $\mathbf{X}_j'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}_j'$ implies that $\mathbf{X}_j'\mathbf{Y}_j = \mathbf{X}_j'\hat{\mathbf{Y}}_j$. Thus, (15-19) can also be written

$$\hat{\boldsymbol{\delta}}_{j,2\text{SLS}} = \begin{bmatrix} \hat{\mathbf{Y}}_j'\hat{\mathbf{Y}}_j & \hat{\mathbf{Y}}_j'\mathbf{X}_j \\ \mathbf{X}_j'\hat{\mathbf{Y}}_j & \mathbf{X}_j'\mathbf{X}_j \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{Y}}_j'\mathbf{y}_j \\ \mathbf{X}_j'\mathbf{y}_j \end{bmatrix}. \tag{15-20}$$

The 2SLS estimator is obtained by ordinary least squares regression of $\mathbf{y}_j$ on $\hat{\mathbf{Y}}_j$ and $\mathbf{X}_j$. Thus, the name stems from the two regressions in the procedure:

1. *Stage 1.* Obtain the least squares predictions from regression of $\mathbf{Y}_j$ on $\mathbf{X}$.
2. *Stage 2.* Estimate $\boldsymbol{\delta}_j$ by least squares regression of $\mathbf{y}_j$ on $\hat{\mathbf{Y}}_j$ and $\mathbf{X}_j$.

A direct proof of the consistency of the 2SLS estimator requires only that we establish that it is a valid IV estimator. For (15-15a), we require

$$\text{plim}\begin{bmatrix} \hat{\mathbf{Y}}_j'\mathbf{Y}_j/T & \hat{\mathbf{Y}}_j'\mathbf{X}_j/T \\ \mathbf{X}_j'\mathbf{Y}_j/T & \mathbf{X}_j'\mathbf{X}_j/T \end{bmatrix} = \text{plim}\begin{bmatrix} \mathbf{P}_j'\mathbf{X}'(\mathbf{X}\boldsymbol{\Pi}_j + \mathbf{V}_j)/T & \mathbf{P}_j'\mathbf{X}'\mathbf{X}_j/T \\ \mathbf{X}_j'(\mathbf{X}\boldsymbol{\Pi}_j + \mathbf{V}_j)/T & \mathbf{X}_j'\mathbf{X}_j/T \end{bmatrix}$$

to be a finite nonsingular matrix. We have used (15-13) for $\mathbf{Y}_j$, which is a continuous function of $\mathbf{P}_j$, which has plim $\mathbf{P}_j = \boldsymbol{\Pi}_j$. The Slutsky theorem thus allows us to substitute $\boldsymbol{\Pi}_j$ for $\mathbf{P}_j$ in the probability limit. That the parts converge to a finite matrix follows from (15-3) and (15-5). It will be nonsingular if $\boldsymbol{\Pi}_j$ has full column rank, which, in turn, will be true if the equation is identified.[15] For (15-15b), we require that

$$\text{plim}\,\frac{1}{T}\begin{bmatrix} \hat{\mathbf{Y}}_j'\boldsymbol{\varepsilon}_j \\ \mathbf{X}_j'\boldsymbol{\varepsilon}_j \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

The second part is assumed in (15-4). For the first, by direct substitution,

$$\text{plim}\,\frac{1}{T}\hat{\mathbf{Y}}_j'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}_j = \text{plim}\left(\frac{\mathbf{Y}_j'\mathbf{X}}{T}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)^{-1}\left(\frac{\mathbf{X}'\boldsymbol{\varepsilon}_j}{T}\right).$$

The third part on the right converges to zero, whereas the other two converge to finite matrices, which confirms the result. Since $\hat{\boldsymbol{\delta}}_{j,2\text{SLS}}$ is an IV estimator, we can just invoke Theorem 5.3 for the asymptotic distribution. A proof of asymptotic efficiency requires the establishment of the benchmark, which we shall do in the discussion of the MLE.

As a final shortcut that is useful for programming purposes, we note that if $\mathbf{X}_j$ is regressed on $\mathbf{X}$, then a perfect fit is obtained, so $\hat{\mathbf{X}}_j = \mathbf{X}_j$. Using the idempotent matrix $(\mathbf{I} - \mathbf{M})$, (15-20) becomes

$$\hat{\boldsymbol{\delta}}_{j,2\text{SLS}} = \begin{bmatrix} \mathbf{Y}_j'(\mathbf{I}-\mathbf{M})\mathbf{Y}_j & \mathbf{Y}_j'(\mathbf{I}-\mathbf{M})\mathbf{X}_j \\ \mathbf{X}_j'(\mathbf{I}-\mathbf{M})\mathbf{Y}_j & \mathbf{X}_j'(\mathbf{I}-\mathbf{M})\mathbf{X}_j \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y}_j'(\mathbf{I}-\mathbf{M})\mathbf{y}_j \\ \mathbf{X}_j'(\mathbf{I}-\mathbf{M})\mathbf{y}_j \end{bmatrix}.$$

Thus,

$$\begin{aligned} \hat{\boldsymbol{\delta}}_{j,2\text{SLS}} &= [\hat{\mathbf{Z}}_j'\hat{\mathbf{Z}}_j]^{-1}\hat{\mathbf{Z}}_j'\mathbf{y}_j \\ &= [(\mathbf{Z}_j'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}_j)]^{-1}(\mathbf{Z}_j'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_j, \end{aligned} \tag{15-21}$$

where all columns of $\hat{\mathbf{Z}}_j'$ are obtained as predictions in a regression of the corresponding

[15]Schmidt (1976, pp. 150–151) provides a proof of this result.

column of $\mathbf{Z}_j$ on $\mathbf{X}$. This equation also results in a useful simplification of the estimated asymptotic covariance matrix,

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

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

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

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

15.5.4 GMM ESTIMATION

The GMM estimator in Section 10.4 is, with a minor change of notation, precisely the set of procedures we have been using here. Using this method, however, will allow us to generalize the covariance structure for the disturbances. We assume that

$$y_{jt} = \mathbf{z}_{jt}'\boldsymbol{\delta}_j + \varepsilon_{jt},$$

where $\mathbf{z}_{jt} = [\mathbf{Y}_{jt}, \mathbf{x}_{jt}]$ (we use the capital $\mathbf{Y}_{jt}$ to denote the L_j included endogenous variables). Thus far, we have assumed that ε_{jt} in the jth equation is neither heteroscedastic nor autocorrelated. There is no need to impose those assumptions at this point. Autocorrelation in the context of a simultaneous equations model is a substantial complication, however. For the present, we will consider the heteroscedastic case only.

The assumptions of the model provide the orthogonality conditions,

$$E[\mathbf{x}_t\varepsilon_{jt}] = E[\mathbf{x}_t(y_{jt} - \mathbf{z}_{jt}'\boldsymbol{\delta}_j)] = \mathbf{0}.$$

If $\mathbf{x}_t$ is taken to be the full set of exogenous variables in the model, then we obtain the criterion for the GMM estimator,

$$\begin{aligned} q &= \left[\frac{\mathbf{e}(\mathbf{z}_t, \boldsymbol{\delta}_j)'\mathbf{X}}{T}\right]\mathbf{W}_{jj}^{-1}\left[\frac{\mathbf{X}'\mathbf{e}(\mathbf{z}_t, \boldsymbol{\delta}_j)}{T}\right] \\ &= \bar{\mathbf{m}}(\boldsymbol{\delta}_j)'\mathbf{W}_{jj}^{-1}\bar{\mathbf{m}}(\boldsymbol{\delta}_j), \end{aligned}$$

where

$$\bar{\mathbf{m}}(\boldsymbol{\delta}_j) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t(y_{jt} - \mathbf{z}_{jt}'\boldsymbol{\delta}_j) \quad \text{and} \quad \mathbf{W}_{jj}^{-1} = \text{the GMM weighting matrix.}$$

Once again, this is precisely the estimator defined in Section 10.4 [see (10-17)]. If the disturbances are assumed to be homoscedastic and nonautocorrelated, then the optimal weighting matrix will be an estimator of the inverse of

$$\begin{aligned} \mathbf{W}_{jj} &= \text{Asy. Var}[\sqrt{T}\,\bar{\mathbf{m}}(\boldsymbol{\delta}_j)] \\ &= \text{plim}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'(y_{jt} - \mathbf{z}_{jt}'\boldsymbol{\delta}_j)^2\right] \\ &= \text{plim}\,\frac{1}{T}\sum_{t=1}^{T}\sigma_{jj}\mathbf{x}_t\mathbf{x}_t' \\ &= \text{plim}\,\frac{\sigma_{jj}(\mathbf{X}'\mathbf{X})}{T} \end{aligned}$$

The constant σ_{jj} is irrelevant to the solution. If we use $(\mathbf{X}'\mathbf{X})^{-1}$ as the weighting matrix, then the GMM estimator that minimizes q is the 2SLS estimator.

The extension that we can obtain here is to allow for heteroscedasticity of unknown form. There is no need to rederive the earlier result. If the disturbances are heteroscedastic, then

$$\mathbf{W}_{jj} = \text{plim}\,\frac{1}{T}\sum_{t=1}^{T}\omega_{jj,t}\mathbf{x}_t\mathbf{x}_t' = \text{plim}\frac{\mathbf{X}'\boldsymbol{\Omega}_{jj}\mathbf{X}}{T}.$$

The weighting matrix can be estimated with White's consistent estimator—see (10-23)—if a consistent estimator of $\boldsymbol{\delta}_j$ is in hand with which to compute the residuals. One is, since 2SLS ignoring the heteroscedasticity is consistent, albeit inefficient. The conclusion then is that under these assumptions, there is a way to improve on 2SLS by adding another step. The name 3SLS is reserved for the systems estimator of this sort. When choosing between 2.5-stage least squares and Davidson and MacKinnon's suggested "heteroscedastic 2SLS, or **H2SLS**," we chose to opt for the latter. The estimator is based on the initial two-stage least squares procedure. Thus,

$$\hat{\boldsymbol{\delta}}_{j,\text{H2SLS}} = [\mathbf{Z}_j'\mathbf{X}(\mathbf{S}_{0,jj})^{-1}\mathbf{X}'\mathbf{Z}_j]^{-1}[\mathbf{Z}_j'\mathbf{X}(\mathbf{S}_{0,jj})^{-1}\mathbf{X}'\mathbf{y}_j],$$

where

$$\mathbf{S}_{0,jj} = \sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'(y_{jt} - \mathbf{z}_{jt}'\hat{\boldsymbol{\delta}}_{j,\text{2SLS}})^2.$$

The asymptotic covariance matrix is estimated with

$$\text{Est.Asy. Var}[\hat{\boldsymbol{\delta}}_{j,\text{H2SLS}}] = [\mathbf{Z}_j'\mathbf{X}(\mathbf{S}_{0,jj})^{-1}\mathbf{X}'\mathbf{Z}_j]^{-1}.$$

Extensions of this estimator were suggested by Cragg (1983) and Cumby, Huizinga, and Obstfeld (1983).

15.5.5 LIMITED INFORMATION MAXIMUM LIKELIHOOD AND THE *K* CLASS OF ESTIMATORS

The **limited information maximum likelihood (LIML) estimator** is based on a single equation under the assumption of normally distributed disturbances; LIML is efficient among single-equation estimators. A full (lengthy) derivation of the log-likelihood is provided in Theil (1971) and Davidson and MacKinnon (1993). We will proceed to the practical aspects of this estimator and refer the reader to these sources for the background formalities. 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. [See Phillips (1983).] The 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

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

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 above.

The LIML, or **least variance ratio** estimator, can be computed as follows.[16] Let

$$\mathbf{W}_j^0 = \mathbf{E}_j^{0\prime}\mathbf{E}_j^0, \tag{15-22}$$

where

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

and

$$\mathbf{E}_j^0 = \mathbf{M}_j\mathbf{Y}_j^0 = [\mathbf{I} - \mathbf{X}_j(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j']\mathbf{Y}_j^0. \tag{15-23}$$

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, 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

$$\mathbf{W}_j^1 = \mathbf{E}_j^{1\prime}\mathbf{E}_j^1 = \mathbf{Y}_j^{0\prime}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}_j^0. \tag{15-24}$$

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

$$\lambda_1 = \text{smallest characteristic root of } \left(\mathbf{W}_j^1\right)^{-1}\mathbf{W}_j^0. \tag{15-25}$$

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,

$$\hat{\boldsymbol{\gamma}}_{j,\text{LIML}} = \left[\mathbf{W}_{jj}^0 - \lambda_1\mathbf{W}_{jj}^1\right]^{-1}\left(\mathbf{w}_j^0 - \lambda_1\mathbf{w}_j^1\right) \tag{15-26}$$

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

[16]The least variance ratio estimator is derived in Johnston (1984). The LIML estimator was derived by Anderson and Rubin (1949, 1950).

estimator.[17] The implication is that with normally distributed disturbances, 2SLS is fully efficient.

The "k class" of estimators is defined by the following form

$$\hat{\boldsymbol{\delta}}_{j,k} = \begin{bmatrix} \mathbf{Y}_j'\mathbf{Y}_j - k\mathbf{V}_j'\mathbf{V}_j & \mathbf{Y}_j'\mathbf{X}_j \\ \mathbf{X}_j'\mathbf{Y}_j & \mathbf{X}_j'\mathbf{X}_j \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y}_j'\mathbf{y}_j - k\mathbf{V}_j'\mathbf{v}_j \\ \mathbf{X}_j'\mathbf{y}_j \end{bmatrix}.$$

We have already considered three members of the class, OLS with $k = 0$, 2SLS with $k = 1$, and, it can be shown, LIML with $k = \lambda_1$. [This last result follows from (15-26).] There have been many other k-class estimators derived; Davidson and MacKinnon (1993, pp. 649–651) and Mariano (2001) give discussion. It has been shown that all members of the k class for which k converges to 1 at a rate faster than $1/\sqrt{n}$ have the same asymptotic distribution as that of the 2SLS estimator that we examined earlier. These are largely of theoretical interest, given the pervasive use of 2SLS or OLS, save for an important consideration. The large- sample properties of all k-class estimator estimators are the same, but the finite-sample properties are possibly very different. Davidson and MacKinnon (1993) and Mariano (1982, 2001) suggest that some evidence favors LIML when the sample size is small or moderate and the number of overidentifying restrictions is relatively large.

15.5.6 TWO-STAGE LEAST SQUARES IN MODELS THAT ARE NONLINEAR IN VARIABLES

The analysis of simultaneous equations becomes considerably more complicated when the equations are nonlinear. Amemiya presents a general treatment of nonlinear models.[18] A case that is broad enough to include many practical applications is the one analyzed by Kelejian (1971),

$$\mathbf{y}_j = \gamma_{1j}\mathbf{f}_{1j}(\mathbf{y}, \mathbf{x}) + \gamma_{2j}\mathbf{f}_{2j}(\mathbf{y}, \mathbf{x}) + \cdots + \mathbf{X}_j\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j,$$ [19]

which is an extension of (7-4). Ordinary least squares will be inconsistent for the same reasons as before, but an IV estimator, if one can be devised, should have the familiar properties. Because of the nonlinearity, it may not be possible to solve for the reduced-form equations (assuming that they exist), $h_{ij}(\mathbf{x}) = E[f_{ij} \mid \mathbf{x}]$. Kelejian shows that 2SLS based on a Taylor series approximation to h_{ij}, using the linear terms, higher powers, and cross-products of the variables in $\mathbf{x}$, will be consistent. The analysis of 2SLS presented earlier then applies to the $\mathbf{Z}_j$ consisting of $[\hat{\mathbf{f}}_{1j}, \hat{\mathbf{f}}_{2j}, \ldots, \mathbf{X}_j]$. [The alternative approach of using fitted values for $\mathbf{y}$ appears to be inconsistent. See Kelejian (1971) and Goldfeld and Quandt (1968).]

In a linear model, if an equation fails the order condition, then it cannot be estimated by 2SLS. This statement is not true of Kelejian's approach, however, since taking higher powers of the regressors creates many more linearly independent instrumental variables. If an equation in a linear model fails the rank condition but not the order

[17] This is proved by showing that both estimators are members of the "k class" of estimators, all of which have the same asymptotic covariance matrix. Details are given in Theil (1971) and Schmidt (1976).

[18] Amemiya (1985, pp. 245–265). See, as well, Wooldridge (2002, ch. 9).

[19] 2SLS for models that are nonlinear in the parameters is discussed in Chapters 10 and 11 in connection with GMM estimators.

condition, then the 2SLS estimates can be computed in a finite sample but will fail to exist asymptotically because $\mathbf{X}\boldsymbol{\Pi}_j$ will have short rank. Unfortunately, to the extent that Kelejian's approximation never exactly equals the true reduced form unless it happens to be the polynomial in $\mathbf{x}$ (unlikely), this built-in control need not be present, even asymptotically. Thus, although the model in Example 15.7 (below) is unidentified, computation of Kelejian's 2SLS estimator appears to be routine.

Example 15.7 A Nonlinear Model of Industry Structure

The following model of industry structure and performance was estimated by Strickland and Weiss (1976). Note that the square of the endogenous variable, C, appears in the first equation.

$$A = \alpha_0 + \alpha_1 M + \alpha_2 Cd + \alpha_3 C + \alpha_4 C^2 + \alpha_5 Gr + \alpha_6 D + \varepsilon_1,$$
$$C = \beta_0 + \beta_1 A + \beta_2 MES + \varepsilon_2,$$
$$M = \gamma_0 + \gamma_1 K + \gamma_2 Gr + \gamma_3 C + \gamma_4 Gd + \gamma_5 A + \gamma_6 MES + \varepsilon_3.$$

S = industry sales
A = advertising/S,
C = concentration,
Cd = consumer demand/S,
MES = efficient scale/S,
M = price cost margin,
D = durable goods industry(0/1),
Gr = industry growth rate,
K = capital stock/S,
Gd = geographic dispersion.

Since the only restrictions are exclusions, we may check identification by the rule rank $[\mathbf{A}_3', \mathbf{A}_5'] = M - 1$ discussed in Section 15.3.1. Identification of the first equation requires

$$[\mathbf{A}_3', \mathbf{A}_5'] = \begin{bmatrix} \beta_2 & 0 & 0 \\ \gamma_6 & \gamma_1 & \gamma_4 \end{bmatrix}$$

to have rank two, which it does unless $\beta_2 = 0$. Thus, the first equation is identified by the presence of the scale variable in the second equation. It is easily seen that the second equation is overidentified. But for the third,

$$[\mathbf{A}_3', \mathbf{A}_5'] = \begin{bmatrix} \alpha_4 & \alpha_2 & \alpha_6 \\ 0 & 0 & 0 \end{bmatrix} (!),$$

which has rank one, not two. The third equation is not identified. It passes the order condition but fails the rank condition. The failure of the third equation is obvious on inspection. There is no variable in the second equation that is not in the third. Nonetheless, it was possible to obtain two stage least squares estimates because of the nonlinearity of the model and the results discussed above.

15.6 SYSTEM METHODS OF ESTIMATION

We may formulate the full system of equations as

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \vdots \\ \boldsymbol{\delta}_M \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix} \quad \textbf{(15-27)}$$

or

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

where

$$E[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}, \quad \text{and} \quad E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \mathbf{X}] = \bar{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} \otimes \mathbf{I} \tag{15-28}$$

[see (14-3).] 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 in the previous chapter 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 in Chapter 14, we use a generalized least squares approach. Thus, assuming that the matrix of instrumental variables, $\bar{\mathbf{W}}$ satisfies the requirements for an IV estimator, a consistent though inefficient estimator would be

$$\hat{\delta}_{\text{IV}} = [\bar{\mathbf{W}}'\mathbf{Z}]^{-1}\bar{\mathbf{W}}'\mathbf{y}. \tag{15-29}$$

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

$$\hat{\delta}_{\text{IV,GLS}} = [\bar{\mathbf{W}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Z}]^{-1}\bar{\mathbf{W}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{y} \tag{15-30}$$

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

$$\hat{\delta}_{\text{IV,GLS}} = \begin{bmatrix} \sigma^{11}\mathbf{W}_1'\mathbf{Z}_1 & \sigma^{12}\mathbf{W}_1'\mathbf{Z}_2 & \cdots & \sigma^{1M}\mathbf{W}_1'\mathbf{Z}_M \\ \sigma^{21}\mathbf{W}_2'\mathbf{Z}_1 & \sigma^{22}\mathbf{W}_2'\mathbf{Z}_2 & \cdots & \sigma^{2M}\mathbf{W}_2'\mathbf{Z}_M \\ & & \vdots & \\ \sigma^{M1}\mathbf{W}_M'\mathbf{Z}_1 & \sigma^{M2}\mathbf{W}_M'\mathbf{Z}_2 & \cdots & \sigma^{MM}\mathbf{W}_M'\mathbf{Z}_M \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{M} \sigma^{1j}\mathbf{W}_1'\mathbf{y}_j \\ \sum_{j=1}^{M} \sigma^{2j}\mathbf{W}_2'\mathbf{y}_j \\ \vdots \\ \sum_{j=1}^{M} \sigma^{Mj}\mathbf{W}_M'\mathbf{y}_j \end{bmatrix}.$$

Three techniques are generally used for joint estimation of the entire system of equations: three-stage least squares, GMM, and full information maximum likelihood.

15.6.1 THREE-STAGE LEAST SQUARES

Consider the IV estimator formed from

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

The IV estimator

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

is simply equation-by-equation 2SLS. We have already established the consistency of 2SLS. By analogy to the seemingly unrelated regressions model of Chapter 14, however, we would expect this estimator to be less efficient than a GLS estimator. A natural candidate would be

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

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

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

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. [But, see Mariano (2001) on the subject of "weak instruments."]

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

$$\hat{\boldsymbol{\delta}}_{3SLS} = [\hat{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\hat{\mathbf{Z}}]^{-1}\hat{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{y}. \tag{15-31}$$

The appropriate asymptotic covariance matrix for the estimator is

$$\text{Asy. Var}[\hat{\boldsymbol{\delta}}_{3SLS}] = [\bar{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\bar{\mathbf{Z}}]^{-1}, \tag{15-32}$$

where $\bar{\mathbf{Z}} = \text{diag}[\mathbf{X}\boldsymbol{\Pi}_j, \mathbf{X}_j]$. This matrix would be estimated with the bracketed inverse matrix in (15-31).

Using sample data, we find that $\bar{\mathbf{Z}}$ may be estimated with $\hat{\mathbf{Z}}$. The remaining difficulty is to obtain an estimate of $\boldsymbol{\Sigma}$. In estimation of the multivariate regression model, for efficient estimation (that remains to be shown), 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 $\hat{\mathbf{Y}}_j$ for each equation.
2. Compute $\hat{\boldsymbol{\delta}}_{j,2SLS}$ for each equation; then

$$\hat{\sigma}_{ij} = \frac{(\mathbf{y}_i - \mathbf{Z}_i\hat{\boldsymbol{\delta}}_i)'(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_j)}{T}. \tag{15-33}$$

3. Compute the GLS estimator according to (15-31) and an estimate of the asymptotic covariance matrix according to (15-32) using $\hat{\mathbf{Z}}$ and $\hat{\boldsymbol{\Sigma}}$.

It is also possible to iterate the 3SLS computation. Unlike the seemingly unrelated regressions estimator, however, this method does not provide the maximum likelihood estimator, nor does it improve the asymptotic efficiency.[20]

[20] A Jacobian term needed to maximize the log-likelihood is not treated by the 3SLS estimator. See Dhrymes (1973).

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 among all IV estimators that use only the sample information embodied in the system, 3SLS is asymptotically efficient.[21] For normally distributed disturbances, it can also be shown that 3SLS has the same asymptotic distribution as the full-information maximum likelihood estimator, which is asymptotically efficient among all estimators. A direct proof based on the information matrix is possible, but we shall take a much simpler route by simply exploiting a handy result due to Hausman in the next section.

15.6.2 FULL-INFORMATION MAXIMUM LIKELIHOOD

Because of their simplicity and asymptotic efficiency, 2SLS and 3SLS are used almost exclusively (when ordinary least squares is not used) for the estimation of simultaneous-equations models. Nonetheless, it is occasionally useful to obtain maximum likelihood estimates directly. The **full-information maximum likelihood (FIML) estimator** is based on the entire system of equations. With normally distributed disturbances, FIML is efficient among all estimators.

The FIML estimator treats all equations and all parameters jointly. To formulate the appropriate log-likelihood function, we begin with the reduced form,

$$\mathbf{Y} = \mathbf{X\Pi} + \mathbf{V},$$

where each row of $\mathbf{V}$ is assumed to be multivariate normally distributed, with $E[\mathbf{v}_t \mid \mathbf{X}] = \mathbf{0}$ and covariance matrix, $E[\mathbf{v}_t\mathbf{v}_t' \mid \mathbf{X}] = \mathbf{\Omega}$. The log-likelihood for this model is precisely that of the seemingly unrelated regressions model of Chapter 14. For the moment, we can ignore the relationship between the structural and reduced-form parameters. Thus, from (14-20),

$$\ln L = -\frac{T}{2}[M\ln(2\pi) + \ln|\mathbf{\Omega}| + \operatorname{tr}(\mathbf{\Omega}^{-1}\mathbf{W})],$$

where

$$\mathbf{W}_{ij} = \frac{1}{T}(\mathbf{y} - \mathbf{X}\boldsymbol{\pi}_i^0)'(\mathbf{y} - \mathbf{X}\boldsymbol{\pi}_j^0)$$

and

$$\boldsymbol{\pi}_j^0 = j\text{th column of } \mathbf{\Pi}.$$

This function is to be maximized subject to all the restrictions imposed by the structure. Make the substitutions $\mathbf{\Pi} = -\mathbf{B\Gamma}^{-1}$ and $\mathbf{\Omega} = (\mathbf{\Gamma}^{-1})'\mathbf{\Sigma\Gamma}^{-1}$ so that $\mathbf{\Omega}^{-1} = \mathbf{\Gamma\Sigma}^{-1}\mathbf{\Gamma}'$. Thus,

$$\ln L = -\frac{T}{2}\left[M\ln(2\pi) + \ln|(\mathbf{\Gamma}^{-1})'\mathbf{\Sigma\Gamma}^{-1}| + \operatorname{tr}\left\{\frac{1}{T}[\mathbf{\Gamma\Sigma}^{-1}\mathbf{\Gamma}'(\mathbf{Y}+\mathbf{XB\Gamma}^{-1})'(\mathbf{Y}+\mathbf{XB\Gamma}^{-1})]\right\}\right],$$

which can be simplified. First,

$$-\frac{T}{2}\ln|(\mathbf{\Gamma}^{-1})'\mathbf{\Sigma\Gamma}^{-1}| = -\frac{T}{2}\ln|\mathbf{\Sigma}| + T\ln|\mathbf{\Gamma}|.$$

[21] See Schmidt (1976) for a proof of its efficiency relative to 2SLS.

Second, $\mathbf{\Gamma}'(\mathbf{Y} + \mathbf{XB}\,\mathbf{\Gamma}^{-1})' = \mathbf{\Gamma}'\mathbf{Y}' + \mathbf{B}'\mathbf{X}'$. By permuting $\mathbf{\Gamma}$ from the beginning to the end of the trace and collecting terms,

$$\text{tr}(\mathbf{\Omega}^{-1}\mathbf{W}) = \text{tr}\left[\frac{\mathbf{\Sigma}^{-1}(\mathbf{Y\Gamma} + \mathbf{XB})'(\mathbf{Y\Gamma} + \mathbf{XB})}{T}\right].$$

Therefore, the log-likelihood is

$$\ln L = -\frac{T}{2}\,[M \ln(2\pi) - 2\ln|\mathbf{\Gamma}| + \text{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}) + \ln|\mathbf{\Sigma}|],$$

where

$$s_{ij} = \frac{1}{T}(\mathbf{Y\Gamma}_i + \mathbf{XB}_i)'(\mathbf{Y\Gamma}_j + \mathbf{XB}_j).$$

[In terms of nonzero parameters, s_{ij} is $\hat{\sigma}_{ij}$ of (15-32).]

In maximizing $\ln L$, it is necessary to impose all the additional restrictions on the structure. The trace may be written in the form

$$\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{M}\sigma^{ij}(\mathbf{y}_i - \mathbf{Y}_i\boldsymbol{\gamma}_i - \mathbf{X}_i\boldsymbol{\beta}_i)'(\mathbf{y}_j - \mathbf{Y}_j\boldsymbol{\gamma}_j - \mathbf{X}_j\boldsymbol{\beta}_j)}{T}. \qquad \textbf{(15-34)}$$

Maximizing $\ln L$ subject to the exclusions in (15-34) and any other restrictions, if necessary, produces the FIML estimator. This has all the desirable asymptotic properties of maximum likelihood estimators and, therefore, is asymptotically efficient among estimators of the simultaneous-equations model. The asymptotic covariance matrix for the FIML estimator is the same as that for the 3SLS estimator.

A useful interpretation of the FIML estimator is provided by Dhrymes (1973, p. 360) and Hausman (1975, 1983). They show that the FIML estimator of $\boldsymbol{\delta}$ is a fixed point in the equation

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

where

$$\hat{\mathbf{Z}}(\hat{\boldsymbol{\delta}})'(\hat{\mathbf{\Sigma}}^{-1} \otimes \mathbf{I}) = \begin{bmatrix} \hat{\sigma}^{11}\hat{\mathbf{Z}}_1' & \hat{\sigma}^{12}\hat{\mathbf{Z}}_1' & \cdots & \hat{\sigma}^{1M}\hat{\mathbf{Z}}_1' \\ \hat{\sigma}^{12}\hat{\mathbf{Z}}_2' & \hat{\sigma}^{22}\hat{\mathbf{Z}}_2' & \cdots & \hat{\sigma}^{2M}\hat{\mathbf{Z}}_2' \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\sigma}^{1M}\hat{\mathbf{Z}}_M' & \hat{\sigma}^{2M}\hat{\mathbf{Z}}_M' & \cdots & \hat{\sigma}^{MM}\hat{\mathbf{Z}}_M' \end{bmatrix} = \hat{\hat{\mathbf{Z}}}'$$

and

$$\hat{\mathbf{Z}}_j = [\mathbf{X}\hat{\mathbf{\Pi}}_j, \mathbf{X}_j].$$

$\hat{\mathbf{\Pi}}$ is computed from the structural estimates:

$$\hat{\mathbf{\Pi}}_j = M_j \text{ columns of } -\hat{\mathbf{B}}\hat{\mathbf{\Gamma}}^{-1}$$

and

$$\hat{\sigma}_{ij} = \frac{1}{T}(\mathbf{y}_i - \mathbf{Z}_i\hat{\boldsymbol{\delta}}_i)'(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_j) \quad \text{and} \quad \hat{\sigma}^{ij} = (\hat{\mathbf{\Sigma}}^{-1})_{ij}.$$

This result implies that the FIML estimator is also an IV estimator. The asymptotic covariance matrix for the FIML estimator follows directly from its form as an IV estimator. Since this matrix is the same as that of the 3SLS estimator, we conclude that with normally distributed disturbances, 3SLS has the same asymptotic distribution as maximum likelihood. The practical usefulness of this important result has not gone unnoticed by practitioners. The 3SLS estimator is far easier to compute than the FIML estimator. The benefit in computational cost comes at no cost in asymptotic efficiency. As always, the small-sample properties remain ambiguous, but by and large, where a systems estimator is used, 3SLS dominates FIML nonetheless.[22] (One reservation arises from the fact that the 3SLS estimator is robust to nonnormality whereas, because of the term $\ln|\mathbf{\Gamma}|$ in the log-likelihood, the FIML estimator is not. In fact, the 3SLS and FIML estimators are usually quite different numerically.)

15.6.3 GMM ESTIMATION

The GMM estimator for a system of equations is described in Section 14.4.3. As in the single-equation case, a minor change in notation produces the estimators of this chapter. As before, we will consider the case of unknown heteroscedasticity only. The extension to autocorrelation is quite complicated. [See Cumby, Huizinga, and Obstfeld (1983).] The orthogonality conditions defined in (14-46) are

$$E[\mathbf{x}_t\varepsilon_{jt}] = E[\mathbf{x}_t(y_{jt} - \mathbf{z}'_{jt}\boldsymbol{\delta}_j)] = \mathbf{0}.$$

If we consider all the equations jointly, then we obtain the criterion for estimation of all the model's parameters,

$$q = \sum_{j=1}^{M}\sum_{l=1}^{M}\left[\frac{\mathbf{e}(\mathbf{z}_t, \boldsymbol{\delta}_j)'\mathbf{X}}{T}\right][\mathbf{W}]^{jl}\left[\frac{\mathbf{X}'\mathbf{e}(\mathbf{z}_t, \boldsymbol{\delta}_l)}{T}\right]$$

$$= \sum_{j=1}^{M}\sum_{l=1}^{M}\bar{\mathbf{m}}(\boldsymbol{\delta}_j)'[\mathbf{W}]^{jl}\bar{\mathbf{m}}(\boldsymbol{\delta}_l),$$

where

$$\bar{\mathbf{m}}(\boldsymbol{\delta}_j) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t(y_{jt} - \mathbf{z}'_{jt}\boldsymbol{\delta}_j)$$

and

$$[\mathbf{W}]^{jl} = \text{block } jl \text{ of the weighting matrix, } \mathbf{W}^{-1}.$$

As before, we consider the optimal weighting matrix obtained as the asymptotic covariance matrix of the empirical moments, $\bar{\mathbf{m}}(\boldsymbol{\delta}_j)$. These moments are stacked in a single vector $\bar{\mathbf{m}}(\boldsymbol{\delta})$. Then, the jlth block of Asy. Var$[\sqrt{T}\,\bar{\mathbf{m}}(\boldsymbol{\delta})]$ is

$$\mathbf{\Phi}_{jl} = \text{plim}\left\{\frac{1}{T}\sum_{t=1}^{T}[\mathbf{x}_t\mathbf{x}'_t(y_{jt} - \mathbf{z}'_{jt}\boldsymbol{\delta}_j)(y_{lt} - \mathbf{z}'_{lt}\boldsymbol{\delta}_l)]\right\} = \text{plim}\left(\frac{1}{T}\sum_{t=1}^{T}\omega_{jl,t}\mathbf{x}_t\mathbf{x}'_t\right).$$

[22] PC-GIVE(8), SAS, and TSP(4.2) are three computer programs that are widely used. A survey is given in Silk (1996).

If the disturbances are homoscedastic, then $\mathbf{\Phi}_{jl} = \sigma_{jl}[\text{plim}(\mathbf{X}'\mathbf{X}/T)]$ is produced. Otherwise, we obtain a matrix of the form $\mathbf{\Phi}_{jl} = \text{plim}[\mathbf{X}'\mathbf{\Omega}_{jl}\mathbf{X}/T]$. Collecting terms, then, the criterion function for GMM estimation is

$$q = \begin{bmatrix} [\mathbf{X}'(\mathbf{y}_1 - \mathbf{Z}_1\boldsymbol{\delta}_1)]/T \\ [\mathbf{X}'(\mathbf{y}_2 - \mathbf{Z}_2\boldsymbol{\delta}_2)]/T \\ \vdots \\ [\mathbf{X}'(\mathbf{y}_M - \mathbf{Z}_M\boldsymbol{\delta}_M)]/T \end{bmatrix}' \begin{bmatrix} \mathbf{\Phi}_{11} & \mathbf{\Phi}_{12} & \cdots & \mathbf{\Phi}_{1M} \\ \mathbf{\Phi}_{21} & \mathbf{\Phi}_{22} & \cdots & \mathbf{\Phi}_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{\Phi}_{M1} & \mathbf{\Phi}_{M2} & \cdots & \mathbf{\Phi}_{MM} \end{bmatrix}^{-1} \begin{bmatrix} [\mathbf{X}'(\mathbf{y}_1 - \mathbf{Z}_1\boldsymbol{\delta}_1)]/T \\ [\mathbf{X}'(\mathbf{y}_2 - \mathbf{Z}_2\boldsymbol{\delta}_2)]/T \\ \vdots \\ [\mathbf{X}'(\mathbf{y}_M - \mathbf{Z}_M\boldsymbol{\delta}_M)]/T \end{bmatrix}.$$

For implementation, $\mathbf{\Phi}_{jl}$ can be estimated with

$$\hat{\mathbf{\Phi}}_{jl} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'(y_{jt} - \mathbf{z}_{jt}'\mathbf{d}_j)(y_{lt} - \mathbf{z}_{lt}'\mathbf{d}_l),$$

where $\mathbf{d}_j$ is a consistent estimator of $\boldsymbol{\delta}_j$. The two-stage least squares estimator is a natural choice. For the diagonal blocks, this choice is the White estimator as usual. For the off-diagonal blocks, it is a simple extension. With this result in hand, the first-order conditions for GMM estimation are

$$\frac{\partial\hat{q}}{\partial\boldsymbol{\delta}_j} = 2\sum_{l=1}^{M}\left(\frac{\mathbf{Z}_j'\mathbf{X}}{T}\right)\hat{\mathbf{\Phi}}^{jl}\left[\frac{\mathbf{X}'(\mathbf{y}_l - \mathbf{Z}_l\boldsymbol{\delta}_l)}{T}\right]$$

where $\hat{\mathbf{\Phi}}^{jl}$ is the jlth block in the inverse of the estimate if the center matrix in q.

The solution is

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

The asymptotic covariance matrix for the estimator would be estimated with T times the large inverse matrix in brackets.

Several of the estimators we have already considered are special cases:

- If $\hat{\mathbf{\Phi}}_{jj} = \hat{\sigma}_{jj}(\mathbf{X}'\mathbf{X}/T)$ and $\hat{\mathbf{\Phi}}_{jl} = \mathbf{0}$ for $j \neq l$, then $\hat{\boldsymbol{\delta}}_j$ is 2SLS.
- If $\hat{\mathbf{\Phi}}_{jl} = \mathbf{0}$ for $j \neq l$, then $\hat{\boldsymbol{\delta}}_j$ is H2SLS, the single-equation GMM estimator.
- If $\hat{\mathbf{\Phi}}_{jl} = \hat{\sigma}_{jl}(\mathbf{X}'\mathbf{X}/T)$, then $\hat{\boldsymbol{\delta}}_j$ is 3SLS.

As before, the GMM estimator brings efficiency gains in the presence of heteroscedasticity. If the disturbances are homoscedastic, then it is asymptotically the same as 3SLS, [although in a finite sample, it will differ numerically because $\mathbf{S}_{jl}$ will not be identical to $\hat{\sigma}_{jl}(\mathbf{X}'\mathbf{X})$].

15.6.4 RECURSIVE SYSTEMS AND EXACTLY IDENTIFIED EQUATIONS

Finally, there are two special cases worth noting. First, for the fully recursive model,

1. $\boldsymbol{\Gamma}$ is upper triangular, with ones on the diagonal. Therefore, $|\boldsymbol{\Gamma}| = 1$ and $\ln|\boldsymbol{\Gamma}| = 0$.
2. $\boldsymbol{\Sigma}$ is diagonal, so $\ln|\boldsymbol{\Sigma}| = \sum_{j=1}^{M} \ln \sigma_{jj}$ and the trace in the exponent becomes

$$\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) = \sum_{j=1}^{M} \frac{1}{\sigma_{jj}} \frac{1}{T} (\mathbf{y}_j - \mathbf{Y}_j\boldsymbol{\gamma}_j - \mathbf{X}_j\boldsymbol{\beta}_j)'(\mathbf{y}_j - \mathbf{Y}_j\boldsymbol{\gamma}_j - \mathbf{X}_j\boldsymbol{\beta}_j).$$

The log-likelihood reduces to $\ln L = \sum_{j=1}^{M} \ln L_j$, where

$$\ln L_j = -\frac{T}{2}[\ln(2\pi) + \ln \sigma_{jj}] - \frac{1}{2\sigma_{jj}} (\mathbf{y}_j - \mathbf{Y}_j\boldsymbol{\gamma}_j - \mathbf{X}_j\boldsymbol{\beta}_j)'(\mathbf{y}_j - \mathbf{Y}_j\boldsymbol{\gamma}_j - \mathbf{X}_j\boldsymbol{\beta}_j).$$

Therefore, the FIML estimator for this model is just equation-by-equation least squares. We found earlier that ordinary least squares was consistent in this setting. We now find that it is asymptotically efficient as well.

The second interesting special case occurs when every equation is exactly identified. In this case, $K_j^* = M_j$ in every equation. It is straightforward to show that in this case, 2SLS = 3SLS = LIML = FIML, and $\hat{\boldsymbol{\delta}}_j = [\mathbf{X}'\mathbf{Z}_j]^{-1}\mathbf{X}'\mathbf{y}_j$.

15.7 COMPARISON OF METHODS—KLEIN'S MODEL I

The preceding has described a large number of estimators for simultaneous-equations models. As an example, Table 15.3 presents limited- and full-information estimates for Klein's Model I based on the original data for 1921 and 1941. The H3SLS estimates for the system were computed in two pairs, (C, I) and (C, W^p), because there were insufficient observations to fit the system as a whole. The first of these are reported for the C equation.[23]

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.[24] But this result must be tempered by the finding that the OLS standard errors are, in all likelihood, not useful for inference purposes.[25] Nonetheless, OLS is a frequently used estimator. Obviously, this discussion

[23]The asymptotic covariance matrix for the LIML estimator will differ from that for the 2SLS estimator in a finite sample because the estimator of σ_{jj} that multiplies the inverse matrix will differ and because in computing the matrix to be inverted, the value of "k" (see the equation after (15-26)) is one for 2SLS and the smallest root in (15-25) for LIML. Asymptotically, k equals one and the estimators of σ_{jj} are equivalent.

[24]See Goldberger (1964, pp. 359–360).

[25]Cragg (1967).

TABLE 15.3 Estimates of Klein's Model I (Estimated Asymptotic Standard Errors in Parentheses)

	Limited-Information Estimates				*Full-Information Estimates*			
	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)	(0.033)
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)	(0.038)
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				**FIML**			
C	17.1	−0.222	0.396	0.823	18.3	−0.232	0.388	0.802
	(1.84)	(0.202)	(0.174)	(0.055)	(2.49)	(0.312)	(0.217)	(0.036)
I	22.6	0.075	0.680	−0.168	27.3	−0.801	1.052	−0.146
	(9.24)	(0.219)	(0.203)	(0.044)	(7.94)	(0.491)	(0.353)	(0.30)
W^p	1.53	0.434	0.151	0.132	5.79	0.234	0.285	0.235
	(2.40)	(0.137)	(0.135)	(0.065)	(1.80)	(0.049)	(0.045)	(0.035)
	GMM (H2SLS)				**GMM (H3SLS)**			
C	14.3	0.090	0.143	0.864	15.7	0.068	0.167	0.829
	(0.897)	(0.062)	(0.065)	(0.029)	(0.951)	(0.091)	(0.080)	(0.033)
I	23.5	0.146	0.591	−0.171	20.6	0.213	−0.520	−0.157
	(6.40)	(0.120)	(0.129)	(0.031)	(4.89)	(0.087)	(0.099)	(0.025)
W^p	3.06	0.455	0.106	0.130	2.09	0.446	0.131	0.112
	(0.64)	(0.028)	(0.030)	(0.022)	(0.510)	(0.019)	(0.021)	(0.021)
	OLS				**I3SLS**			
C	16.2	0.193	0.090	0.796	16.6	0.165	0.177	0.766
	(1.30)	(0.091)	(0.091)	(0.040)	(1.22)	(0.096)	(0.090)	(0.035)
I	10.1	0.480	0.333	−0.112	42.9	−0.356	1.01	−0.260
	(5.47)	(0.097)	(0.101)	(0.027)	(10.6)	(0.260)	(0.249)	(0.051)
W^p	1.50	0.439	0.146	0.130	2.62	0.375	0.194	0.168
	(1.27)	(0.032)	(0.037)	(0.032)	(1.20)	(0.031)	(0.032)	(0.029)

is relevant only to finite samples. Asymptotically, 2SLS must dominate OLS, and in a correctly specified model, any full-information estimator must dominate any limited-information one. 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.

The large difference between the inconsistent OLS and the other estimates suggests the bias discussed earlier. On the other hand, the incorrect sign on the LIML and FIML estimate of the coefficient on P and the even larger difference of the coefficient on P_{-1} in the C equation are striking. Assuming that the equation is properly specified, these anomalies would likewise be attributed to finite sample variation, because LIML and 2SLS are asymptotically equivalent. The GMM estimator is also striking. The estimated standard errors are noticeably smaller for all the coefficients. It should be noted, however, that this estimator is based on a presumption of heteroscedasticity when in this time series, there is little evidence of its presence. The results are broadly suggestive, but the appearance of having achieved something for nothing is deceiving. Our earlier results on the efficiency of 2SLS are intact. If there is heteroscedasticity, then 2SLS is no longer fully efficient, but, then again, neither is H2SLS. The latter is more efficient than the former in the presence of heteroscedasticity, but it is equivalent to 2SLS in its absence.

Intuition would suggest that systems methods, 3SLS, GMM, and FIML, are to be preferred to single-equation methods, 2SLS and LIML. Indeed, since the advantage is so transparent, why would one ever choose a single-equation estimator? The proper analogy is to the use of single-equation OLS versus GLS in the SURE model of Chapter 14. An obvious practical consideration is the computational simplicity of the single-equation methods. But the current state of available software has all but eliminated this advantage.

Although the systems methods 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 or FIML. The limited-information estimators will, by and large, confine a problem to the particular equation in which it appears. Second, in the same fashion as the SURE 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. Although they are only single estimates, the results for Klein's Model I give a striking example. The upshot would appear to be that the advantage of the systems estimators in finite samples may be more modest than the asymptotic results would suggest. Monte Carlo studies of the issue have tended to reach the same conclusion.[26]

15.8 SPECIFICATION TESTS

In a strident criticism of structural estimation, Liu (1960) argued that all simultaneous-equations models of the economy were truly unidentified and that only reduced forms could be estimated. Although his criticisms may have been exaggerated (and never gained wide acceptance), modelers have been interested in testing the restrictions that overidentify an econometric model.

The first procedure for testing the overidentifying restrictions in a model was developed by Anderson and Rubin (1950). Their likelihood ratio test statistic is a by-product of LIML estimation:

$$\text{LR} = \chi^2[K_j^* - M_j] = T(\lambda_j - 1),$$

where λ_j is the root used to find the LIML estimator. [See (15-27).] The statistic has a limiting chi-squared distribution with degrees of freedom equal to the number of overidentifying restrictions. A large value is taken as evidence that there are exogenous variables in the model that have been inappropriately omitted from the equation being examined. If the equation is exactly identified, then $K_j^* - M_j = 0$, but at the same time, the root will be 1. An alternative based on the Lagrange multiplier principle was

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

proposed by Hausman (1983, p. 433). Operationally, the test requires only the calculation of TR^2, where the R^2 is the uncentered R^2 in the regression of $\hat{\boldsymbol{\varepsilon}}_j = \mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_j$ on all the predetermined variables in the model. The estimated parameters may be computed using 2SLS, LIML, or any other *efficient* limited-information estimator. The statistic has a limiting chi-squared distribution with $K_j^* - M_j$ degrees of freedom under the assumed specification of the model.

Another specification error occurs if the variables assumed to be exogenous in the system are, in fact, correlated with the structural disturbances. Since all the asymptotic properties claimed earlier rest on this assumption, this specification error would be quite serious. Several authors have studied this issue.[27] The specification test devised by Hausman that we used in Section 5.5 in the errors in variables model provides a method of testing for exogeneity in a simultaneous-equations model. Suppose that the variable x^e is in question. The test is based on the existence of two estimators, say $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{\delta}}^*$, such that

under H_0: (x^e is exogenous), both $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{\delta}}^*$ are consistent and $\hat{\boldsymbol{\delta}}^*$ is asymptotically efficient,
under H_1: (x^e is endogenous), $\hat{\boldsymbol{\delta}}$ is consistent, but $\hat{\boldsymbol{\delta}}^*$ is inconsistent.

Hausman bases his version of the test on $\hat{\boldsymbol{\delta}}$ being the 2SLS estimator and $\hat{\boldsymbol{\delta}}^*$ being the 3SLS estimator. A shortcoming of the procedure is that it requires an arbitrary choice of some equation that does not contain x^e for the test. For instance, consider the exogeneity of X_{-1} in the third equation of Klein's Model I. To apply this test, we must use one of the other two equations.

A single-equation version of the test has been devised by Spencer and Berk (1981). We suppose that x^e appears in equation j, so that

$$\begin{aligned}\mathbf{y}_j &= \mathbf{Y}_j\boldsymbol{\gamma}_j + \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{x}^e\theta + \boldsymbol{\varepsilon}_j \\ &= [\mathbf{Y}_j, \mathbf{X}_j, \mathbf{x}^e]\boldsymbol{\delta}_j + \boldsymbol{\varepsilon}_j.\end{aligned}$$

Then $\hat{\boldsymbol{\delta}}^*$ is the 2SLS estimator, treating x^e as an exogenous variable in the system, whereas $\hat{\boldsymbol{\delta}}$ is the IV estimator based on regressing $\mathbf{y}_j$ on $\mathbf{Y}_j, \mathbf{X}_j, \hat{\mathbf{x}}^e$, where the least squares fitted values are based on all the remaining exogenous variables, excluding $\mathbf{x}^e$. The test statistic is then

$$w = (\hat{\boldsymbol{\delta}}^* - \hat{\boldsymbol{\delta}})'\left\{\text{Est. Var}[\hat{\boldsymbol{\delta}}] - \text{Est. Var}[\hat{\boldsymbol{\delta}}^*]\right\}^{-1}(\hat{\boldsymbol{\delta}}^* - \hat{\boldsymbol{\delta}}), \qquad \textbf{(15-35)}$$

which is the Wald statistic based on the difference of the two estimators. The statistic has one degree of freedom. (The extension to a set of variables is direct.)

Example 15.8 Testing Overidentifying Restrictions

For Klein's Model I, the test statistics and critical values for the chi-squared distribution for the overidentifying restrictions for the three equations are given in Table 15.4. There are 20 observations used to estimate the model and eight predetermined variables. The overidentifying restrictions for the wage equation are rejected by both single-equation tests. There are two possibilities. The equation may well be misspecified. Or, as Liu suggests, in a

[27] Wu (1973), Durbin (1954), Hausman (1978), Nakamura and Nakamura (1981) and Dhrymes (1994).

TABLE 15.4 Test Statistics and Critical Values

	λ	LR	TR^2	$K_j^* - M_j$	Chi-Squared Critical Values	$\chi^2[2]$	$\chi^2[3]$
Consumption	1.499	9.98	8.77	2			
Investment	1.086	1.72	1.81	3	5%	5.99	7.82
Wages	2.466	29.3	12.49	3	1%	9.21	11.34

dynamic model, if there is autocorrelation of the disturbances, then the treatment of lagged endogenous variables as if they were exogenous is a specification error.

The results above suggest a specification problem in the third equation of Klein's Model I. To pursue that finding, we now apply the preceding to test the exogeneity of X_{-1}. The two estimated parameter vectors are

$$\hat{\delta}^* = [1.5003, 0.43886, 0.14667, 0.13040] \text{ (i.e., 2SLS)}$$

and

$$\hat{\delta} = [1.2524, 0.42277, 0.167614, 0.13062].$$

Using the Wald criterion, the chi-squared statistic is 1.3977. Thus, the hypothesis (such as it is) is not rejected.

15.9 PROPERTIES OF DYNAMIC MODELS

In models with lagged endogenous variables, the entire previous time path of the exogenous variables and disturbances, not just their current values, determines the current value of the endogenous variables. The intrinsic dynamic properties of the autoregressive model, such as stability and the existence of an equilibrium value, are embodied in their autoregressive parameters. In this section, we are interested in long- and short-run multipliers, stability properties, and simulated time paths of the dependent variables.

15.9.1 DYNAMIC MODELS AND THEIR MULTIPLIERS

The structural form of a dynamic model is

$$\mathbf{y}_t'\mathbf{\Gamma} + \mathbf{x}_t'\mathbf{B} + \mathbf{y}_{t-1}'\mathbf{\Phi} = \boldsymbol{\varepsilon}_t'. \tag{15-36}$$

If the model contains additional lags, then we can add additional equations to the system of the form $\mathbf{y}_{t-1}' = \mathbf{y}_{t-1}'$. For example, a model with two periods of lags would be written

$$[\mathbf{y}_t \quad \mathbf{y}_{t-1}]' \begin{bmatrix} \mathbf{\Gamma} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} + \mathbf{x}_t'[\mathbf{B} \quad \mathbf{0}] + [\mathbf{y}_{t-1} \quad \mathbf{y}_{t-2}]' \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{I} \\ \mathbf{\Phi}_2 & \mathbf{0} \end{bmatrix} = [\boldsymbol{\varepsilon}_t' \quad \mathbf{0}']$$

which can be treated as a model with only a single lag—this is in the form of (15-36). The reduced form is

$$\mathbf{y}_t' = \mathbf{x}_t'\mathbf{\Pi} + \mathbf{y}_{t-1}'\mathbf{\Delta} + \mathbf{v}_t',$$

where

$$\mathbf{\Pi} = -\mathbf{B}\mathbf{\Gamma}^{-1}$$

and

$$\mathbf{\Delta} = -\mathbf{\Phi}\mathbf{\Gamma}^{-1}.$$

From the reduced form,

$$\frac{\partial y_{t,m}}{\partial x_{t,k}} = \mathbf{\Pi}_{km}.$$

The short-run effects are the coefficients on the current xs, so $\mathbf{\Pi}$ is the matrix of **impact multipliers.** By substituting for $\mathbf{y}_{t-1}$ in (15-36), we obtain

$$\mathbf{y}_t' = \mathbf{x}_t'\mathbf{\Pi} + \mathbf{x}_{t-1}'\mathbf{\Pi}\mathbf{\Delta} + \mathbf{y}_{t-2}'\mathbf{\Delta}^2 + (\mathbf{v}_t' + \mathbf{v}_{t-1}'\mathbf{\Delta}).$$

(This manipulation can easily be done with the lag operator—see Section 19.2.2—but it is just as convenient to proceed in this fashion for the present.) Continuing this method for the full t periods, we obtain

$$\mathbf{y}_t' = \sum_{s=0}^{t-1} [\mathbf{x}_{t-s}'\mathbf{\Pi}\mathbf{\Delta}^s] + \mathbf{y}_0'\mathbf{\Delta}^t + \sum_{s=0}^{t-1} \mathbf{v}_{t-s}'\mathbf{\Delta}^s. \qquad \textbf{(15-37)}$$

This shows how the **initial conditions** $\mathbf{y}_0$ and the subsequent time path of the exogenous variables and disturbances completely determine the current values of the endogenous variables. The coefficient matrices in the bracketed sum are the **dynamic multipliers,**

$$\frac{\partial y_{t,m}}{\partial x_{t-s,k}} = (\mathbf{\Pi}\mathbf{\Delta}^s)_{km}.$$

The **cumulated multipliers** are obtained by adding the matrices of dynamic multipliers. If we let s go to infinity in (15-37), then we obtain the **final form** of the model,[28]

$$\mathbf{y}_t' = \sum_{s=0}^{\infty} [\mathbf{x}_{t-s}'\mathbf{\Pi}\mathbf{\Delta}^s] + \sum_{s=0}^{\infty} [\mathbf{v}_{t-s}'\mathbf{\Delta}^s].$$

Assume for the present that $\lim_{t\to\infty} \mathbf{\Delta}^t = \mathbf{0}$. (This says that $\mathbf{\Delta}$ is nilpotent.) Then the matrix of cumulated multipliers in the final form is

$$\mathbf{\Pi}[\mathbf{I} + \mathbf{\Delta} + \mathbf{\Delta}^2 + \cdots] = \mathbf{\Pi}[\mathbf{I} - \mathbf{\Delta}]^{-1}.$$

These coefficient matrices are the long-run or **equilibrium multipliers.** We can also obtain the cumulated multipliers for s periods as

$$\text{cumulated multipliers} = \mathbf{\Pi}[\mathbf{I} - \mathbf{\Delta}]^{-1}[\mathbf{I} - \mathbf{\Delta}^s].$$

Suppose that the values of $\mathbf{x}$ were permanently fixed at $\bar{\mathbf{x}}$. Then the final form shows that if there are no disturbances, the equilibrium value of $\mathbf{y}_t$ would be

$$\bar{\mathbf{y}}' = \sum_{s=0}^{\infty} [\bar{\mathbf{x}}'\mathbf{\Pi}\mathbf{\Delta}^s] = \bar{\mathbf{x}}' \sum_{s=0}^{\infty} \mathbf{\Pi}\mathbf{\Delta}^s = \bar{\mathbf{x}}'\mathbf{\Pi}[\mathbf{I} - \mathbf{\Delta}]^{-1}. \qquad \textbf{(15-38)}$$

[28]In some treatments, (15-37) is labeled the final form instead. Both forms eliminate the lagged values of the dependent variables from the current value. The dependence of the first form on the initial values may make it simpler to interpret than the second form.

Therefore, the equilibrium multipliers are

$$\frac{\partial \bar{y}_m}{\partial \bar{x}_k} = [\mathbf{\Pi}(\mathbf{I} - \mathbf{\Delta})^{-1}]_{km}.$$

Some examples are shown below for Klein's Model I.

15.9.2 STABILITY

It remains to be shown that the matrix of multipliers in the final form converges. For the analysis to proceed, it is necessary for the matrix $\mathbf{\Delta}^t$ to converge to a zero matrix. Although $\mathbf{\Delta}$ is not a symmetric matrix, it will still have a spectral decomposition of the form

$$\mathbf{\Delta} = \mathbf{C}\mathbf{\Lambda}\mathbf{C}^{-1}, \tag{15-39}$$

where $\mathbf{\Lambda}$ is a diagonal matrix containing the characteristic roots of $\mathbf{\Delta}$ and each column of $\mathbf{C}$ is a right characteristic vector,

$$\mathbf{\Delta}\mathbf{c}_m = \lambda_m \mathbf{c}_m. \tag{15-40}$$

Since $\mathbf{\Delta}$ is not symmetric, the elements of $\mathbf{\Lambda}$ (and $\mathbf{C}$) may be complex. Nonetheless, (A-105) continues to hold:

$$\mathbf{\Delta}^2 = \mathbf{C}\mathbf{\Lambda}\mathbf{C}^{-1}\mathbf{C}\mathbf{\Lambda}\mathbf{C}^{-1} = \mathbf{C}\mathbf{\Lambda}^2\mathbf{C}^{-1} \tag{15-41}$$

and

$$\mathbf{\Delta}^t = \mathbf{C}\mathbf{\Lambda}^t\mathbf{C}^{-1}.$$

It is apparent that whether or not $\mathbf{\Delta}^t$ vanishes as $t \to \infty$ depends on its characteristic roots. The condition is $|\lambda_m| < 1$. For the case of a complex root, $|\lambda_m| = |a + bi| = \sqrt{a^2 + b^2}$. For a given model, the stability may be established by examining the largest or **dominant root.**

With many endogenous variables in the model but only a few lagged variables, $\mathbf{\Delta}$ is a large but sparse matrix. Finding the characteristic roots of large, asymmetric matrices is a rather complex computation problem (although there exists specialized software for doing so). There is a way to make the problem a bit more compact. In the context of an example, in Klein's Model I, $\mathbf{\Delta}$ is 6×6, but with three rows of zeros, it has only rank three and three nonzero roots. (See Table 15.5 in Example 15.9 following.) The following partitioning is useful. Let $\mathbf{y}_{t1}$ be the set of endogenous variables that appear in both current and lagged form, and let $\mathbf{y}_{t2}$ be those that appear only in current form. Then the model may be written

$$[\mathbf{y}'_{t1} \quad \mathbf{y}'_{t2}] = \mathbf{x}'_t[\mathbf{\Pi}_1 \quad \mathbf{\Pi}_2] + [\mathbf{y}'_{t-1,1} \quad \mathbf{y}'_{t-1,2}]\begin{bmatrix} \mathbf{\Delta}_1 & \mathbf{\Delta}_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + [\mathbf{v}'_{t1} \quad \mathbf{v}'_{t2}]. \tag{15-42}$$

The characteristic roots of $\mathbf{\Delta}$ are defined by the characteristic polynomial, $|\mathbf{\Delta} - \lambda\mathbf{I}| = 0$. For the partitioned model, this result is

$$\begin{vmatrix} \mathbf{\Delta}_1 - \lambda\mathbf{I} & \mathbf{\Delta}_2 \\ 0 & -\lambda\mathbf{I} \end{vmatrix} = 0.$$

We may use (A-72) to obtain

$$|\boldsymbol{\Delta} - \lambda\mathbf{I}| = (-\lambda)^{M_2}|\boldsymbol{\Delta}_1 - \lambda\mathbf{I}| = 0,$$

where M_2 is the number of variables in $\mathbf{y}_2$. Consequently, we need only concern ourselves with the submatrix of $\boldsymbol{\Delta}$ that defines explicit autoregressions. The part of the reduced form defined by $\mathbf{y}'_{t2} = \mathbf{x}'_t\boldsymbol{\Pi}_2 + \mathbf{y}'_{t-1,1}\boldsymbol{\Delta}_2$ is not directly relevant.

15.9.3 ADJUSTMENT TO EQUILIBRIUM

The adjustment of a dynamic model to an equilibrium involves the following conceptual experiment. We assume that the exogenous variables $\mathbf{x}_t$ have been fixed at a level $\bar{\mathbf{x}}$ for a long enough time that the endogenous variables have fully adjusted to their equilibrium $\bar{\mathbf{y}}$ [defined in (15-38)]. In some arbitrarily chosen period, labeled period 0, an exogenous one-time shock hits the system, so that in period $t = 0$, $\mathbf{x}_t = \mathbf{x}_0 \neq \bar{\mathbf{x}}$. Thereafter, $\mathbf{x}_t$ returns to its former value $\bar{\mathbf{x}}$, and $\mathbf{x}_t = \bar{\mathbf{x}}$ for all $t > 0$. We know from the expression for the final form that, if disturbed, $\mathbf{y}_t$ will ultimately return to the equilibrium. That situation is ensured by the stability condition. Here we consider the time path of the adjustment. Since our only concern at this point is with the exogenous shock, we will ignore the disturbances in the analysis.

At time 0, $\mathbf{y}'_0 = \mathbf{x}'_0\boldsymbol{\Pi} + \mathbf{y}'_{-1}\boldsymbol{\Delta}$. But prior to time 0, the system was in equilibrium, so $\mathbf{y}'_0 = \mathbf{x}'_0\boldsymbol{\Pi} + \bar{\mathbf{y}}'\boldsymbol{\Delta}$. The initial displacement due to the shock to $\bar{\mathbf{x}}$ is

$$\mathbf{y}'_0 - \bar{\mathbf{y}}' = \mathbf{x}'_0\boldsymbol{\Pi} - \bar{\mathbf{y}}'(\mathbf{I} - \boldsymbol{\Delta}).$$

Substituting $\bar{\mathbf{x}}'\boldsymbol{\Pi} = \bar{\mathbf{y}}'(\mathbf{I} - \boldsymbol{\Delta})$ produces

$$\mathbf{y}'_0 - \bar{\mathbf{y}}' = (\mathbf{x}'_0 - \bar{\mathbf{x}}')\boldsymbol{\Pi}. \tag{15-43}$$

As might be expected, the initial displacement is determined entirely by the exogenous shock occurring in that period. Since $\mathbf{x}_t = \bar{\mathbf{x}}$ after period 0, (15-37) implies that

$$\begin{aligned}
\mathbf{y}'_t &= \sum_{s=0}^{t-1} \bar{\mathbf{x}}'\boldsymbol{\Pi}\boldsymbol{\Delta}^s + \mathbf{y}'_0\boldsymbol{\Delta}^t \\
&= \bar{\mathbf{x}}'\boldsymbol{\Pi}(\mathbf{I} - \boldsymbol{\Delta})^{-1}(\mathbf{I} - \boldsymbol{\Delta}^t) + \mathbf{y}'_0\boldsymbol{\Delta}^t \\
&= \bar{\mathbf{y}}' - \bar{\mathbf{y}}'\boldsymbol{\Delta}^t + \mathbf{y}'_0\boldsymbol{\Delta}^t \\
&= \bar{\mathbf{y}}' + (\mathbf{y}'_0 - \bar{\mathbf{y}}')\boldsymbol{\Delta}^t.
\end{aligned}$$

Thus, the entire time path is a function of the initial displacement. By inserting (15-43), we see that

$$\mathbf{y}'_t = \bar{\mathbf{y}}' + (\mathbf{x}'_0 - \bar{\mathbf{x}}')\boldsymbol{\Pi}\boldsymbol{\Delta}^t. \tag{15-44}$$

Since $\lim_{t\to\infty}\boldsymbol{\Delta}^t = \mathbf{0}$, the path back to the equilibrium subsequent to the exogenous shock $(\mathbf{x}_0 - \bar{\mathbf{x}})$ is defined. The stability condition imposed on $\boldsymbol{\Delta}$ ensures that if the system is disturbed at some point by a one-time shock, then barring further shocks or disturbances, it will return to its equilibrium. Since $\mathbf{y}_0, \bar{\mathbf{x}}, \mathbf{x}_0$, and $\mathbf{\Pi}$ are fixed for all time, the shape of the path is completely determined by the behavior of $\mathbf{\Delta}^t$, which we now examine.

In the preceding section, in (15-39) to (15-42), we used the characteristic roots of $\mathbf{\Delta}$ to infer the (lack of) stability of the model. The spectral decomposition of $\mathbf{\Delta}^t$ given in (15-41) may be written

$$\mathbf{\Delta}^t = \sum_{m=1}^{M} \lambda_m^t \mathbf{c}_m \mathbf{d}_m',$$

where $\mathbf{c}_m$ is the mth column of $\mathbf{C}$ and $\mathbf{d}_m'$ is the mth row of $\mathbf{C}^{-1}$.[29] Inserting this result in (15-44), gives

$$\begin{aligned}(\mathbf{y}_t - \bar{\mathbf{y}})' &= [(\mathbf{x}_0 - \bar{\mathbf{x}})'\mathbf{\Pi}] \sum_{m=1}^{M} \lambda_m^t \mathbf{c}_m \mathbf{d}_m' \\ &= \sum_{m=1}^{M} \lambda_m^t [(\mathbf{x}_0 - \bar{\mathbf{x}})'\mathbf{\Pi}\mathbf{c}_m \mathbf{d}_m'] = \sum_{m=1}^{M} \lambda_m^t \mathbf{g}_m'.\end{aligned}$$

(Note that this equation may involve fewer than M terms, since some of the roots may be zero. For Klein's Model I, $M = 6$, but there are only three nonzero roots.) Since $\mathbf{g}_m$ depends only on the initial conditions and the parameters of the model, the behavior of the time path of $(\mathbf{y}_t - \bar{\mathbf{y}})$ is completely determined by λ_m^t. In each period, the deviation from the equilibrium is a sum of M terms of powers of λ_m times a constant. (Each variable has its own set of constants.) The terms in the sum behave as follows:

$$\begin{aligned}&\lambda_m \text{ real} > 0, \quad \lambda_m^t \text{ adds a damped exponential term,}\\ &\lambda_m \text{ real} < 0, \quad \lambda_m^t \text{ adds a damped sawtooth term,}\\ &\lambda_m \text{ complex,} \quad \lambda_m^t \text{ adds a damped sinusoidal term.}\end{aligned}$$

If we write the complex root $\lambda_m = a + bi$ in polar form, then $\lambda = A[\cos B + i \sin B]$, where $A = [a^2 + b^2]^{1/2}$ and $B = \arccos(a/A)$ (in radians), the sinusoidal components each have amplitude A^t and period $2\pi/B$.[30]

Example 15.9 Dynamic Model

The 2SLS estimates of the structure and reduced form of Klein's Model I are given in Table 15.5. (Only the nonzero rows of $\hat{\mathbf{\Phi}}$ and $\hat{\mathbf{\Delta}}$ are shown.)

For the 2SLS estimates of Klein's Model I, the relevant submatrix of $\hat{\mathbf{\Delta}}$ is

$$\hat{\mathbf{\Delta}}_1 = \begin{array}{c} \begin{matrix} K & P & K \end{matrix} \\ \begin{bmatrix} 0.172 & -0.051 & -0.008 \\ 1.511 & 0.848 & 0.743 \\ -0.287 & -0.161 & 0.818 \end{bmatrix} \end{array} \begin{matrix} X_{-1} \\ P_{-1}. \\ K_{-1} \end{matrix}$$

[29] See Section A.6.9.

[30] Goldberger (1964, p. 378).

TABLE 15.5 2SLS Estimates of Coefficient Matrices in Klein's Model I

Variable		*Equation* C	I	W^p	X	P	K
	C	1	0	0	−1	0	0
	I	0	1	0	−1	0	−1
$\hat{\mathbf{\Gamma}} =$	W^p	−0.810	0	1	0	1	0
	X	0	0	−0.439	1	−1	0
	P	−0.017	−0.15	0	0	1	0
	K	0	0	0	0	0	1
	1	−16.555	−20.278	−1.5	0	0	0
	W^g	−0.810	0	0	0	0	0
$\hat{\mathbf{B}} =$	T	0	0	0	0	1	0
	G	0	0	0	−1	0	0
	A	0	0	−0.13	0	0	0
	X_{-1}	0	0	−0.147	0	0	0
$\hat{\mathbf{\Phi}} =$	P_{-1}	−0.216	−0.6160	0	0	0	0
	K_{-1}	0	0.158	0	0	0	−1
	1	42.80	25.83	31.63	68.63	37.00	25.83
	W^g	1.35	0.124	0.646	1.47	0.825	0.125
$\hat{\mathbf{\Pi}} =$	T	−0.128	−0.176	−0.133	−0.303	−1.17	−0.176
	G	0.663	0.153	0.797	1.82	1.02	0.153
	A	0.159	−0.007	0.197	0.152	−0.045	−0.007
	X_{-1}	0.179	−0.008	0.222	0.172	−0.051	−0.008
$\hat{\mathbf{\Delta}} =$	P_{-1}	0.767	0.743	0.663	1.511	0.848	0.743
	K_{-1}	−0.105	−0.182	−0.125	−0.287	−0.161	0.818

The characteristic roots of this matrix are 0.2995 and the complex pair $0.7692 \pm 0.3494i = 0.8448\,[\cos 0.4263 \pm i \sin 0.4263]$. The moduli of the complex roots are 0.8448, so we conclude that the model is stable. The period for the oscillations is $2\pi/0.4263 = 14.73$ periods (years). (See Figure 15.2.)

For a particular variable or group of variables, the various multipliers are submatrices of the multiplier matrices. The dynamic multipliers based on the estimates in Table 15.5 for the effects of the policy variables T and G on output, X, are plotted in Figure 15.2 for current and 20 lagged values. A plot of the period multipliers against the lag length is called the **impulse response function.** The policy effects on output are shown in Figure 15.2. The damped sine wave pattern is characteristic of a dynamic system with imaginary roots. When the roots are real, the impulse response function is a monotonically declining function, instead.

This model has the interesting feature that the long-run multipliers of both policy variables for investment are zero. This is intrinsic to the model. The estimated long-run *balanced-budget multiplier* for equal increases in spending and taxes is $2.10 + (-1.48) = 0.62$.

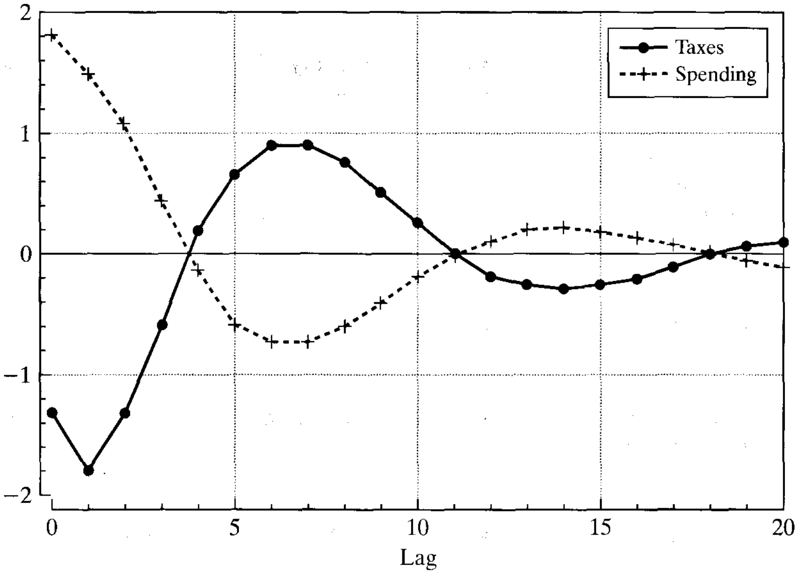

FIGURE 15.2 Impulse Response Function.

15.10 SUMMARY AND CONCLUSIONS

The 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 a number of single equation techniques including least squares, instrumental variables, GMM, 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, GMM, 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. In the final section of this chapter, we examined the special properties of dynamic models. An important consideration in this analysis was the stability of the equations. Modern macroeconometrics involves many models in which one or more roots of the dynamic system equal one, so that these models, in the simple autoregressive form are unstable. In terms of the analysis in Section 15.9.3, in such a model, a shock to the system is permanent—the effects do not die out. We will examine a model of monetary policy with these characteristics in Example 19.6.8.

Key Terms and Concepts

- Admissible
- Behavioral equation
- Causality
- Complete system
- Completeness condition
- Consistent estimates
- Cumulative multiplier
- Dominant root
- Dynamic model
- Dynamic multiplier
- Econometric model
- Endogenous
- Equilibrium condition
- Equilibrium multipliers
- Exactly identified model
- Exclusion restrictions
- Exogenous
- FIML
- Final form
- Full information
- Fully recursive model
- GMM estimation
- Granger causality
- Identification
- Impact multiplier
- Impulse response function
- Indirect least squares
- Initial conditions
- Instrumental variable estimator
- Interdependent
- Jointly dependent
- k class
- Least variance ratio
- Limited information
- LIML
- Nonlinear system
- Nonsample information
- Nonstructural
- Normalization
- Observationally equivalent
- Order condition
- Overidentification
- Predetermined variable
- Problem of identification
- Rank condition
- Recursive model
- Reduced form
- Reduced-form disturbance
- Restrictions
- Simultaneous-equations bias
- Specification test
- Stability
- Structural disturbance
- Structural equation
- System methods of estimation
- Three-stage least squares
- Triangular system
- Two-stage least squares
- Weakly exogenous

Exercises

1. 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) $\sigma_{12} = 0, \beta_{11} = \beta_{21} = \beta_{22} = \beta_{31} = \beta_{32} = 0$.

2. Verify the rank and order conditions for identification of the second and third behavioral equations in Klein's Model I.

3. Check the identifiability of the parameters of the following model:

$$[y_1 \quad y_2 \quad y_3 \quad y_4]\begin{bmatrix} 1 & \gamma_{12} & 0 & 0 \\ \gamma_{21} & 1 & \gamma_{23} & \gamma_{24} \\ 0 & \gamma_{32} & 1 & \gamma_{34} \\ \gamma_{41} & \gamma_{42} & 0 & 1 \end{bmatrix}$$

$$+[x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5]\begin{bmatrix} 0 & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & 1 & 0 & \beta_{24} \\ \beta_{31} & \beta_{32} & \beta_{33} & 0 \\ 0 & 0 & \beta_{43} & \beta_{44} \\ 0 & \beta_{52} & 0 & 0 \end{bmatrix} + [\varepsilon_1 \quad \varepsilon_2 \quad \varepsilon_3 \quad \varepsilon_4].$$

4. Obtain the reduced form for the model in Exercise 1 under each of the assumptions made in parts a and in parts b1 and b9.
5. 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.

6. 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.

7. An updated version of Klein's Model I was estimated. The relevant submatrix of $\boldsymbol{\Delta}$ is

$$\boldsymbol{\Delta}_1 = \begin{bmatrix} -0.1899 & -0.9471 & -0.8991 \\ 0 & 0.9287 & 0 \\ -0.0656 & -0.0791 & 0.0952 \end{bmatrix}.$$

Is the model stable?

8. Prove that

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

9. Prove that an underidentified equation cannot be estimated by 2SLS.