10-30

This is the translog cost function. If δ_{ij} equals zero, then it reduces to the Cobb–Douglas function we looked at earlier. The cost shares are given by

$$
\begin{aligned}
s_1 &= \frac{\partial \ln c}{\partial \ln p_1} = \beta_1 + \delta_{11}\ln p_1 + \delta_{12}\ln p_2 + \cdots + \delta_{1M}\ln p_M,\\
s_2 &= \frac{\partial \ln c}{\partial \ln p_2} = \beta_2 + \delta_{21}\ln p_1 + \delta_{22}\ln p_2 + \cdots + \delta_{2M}\ln p_M,\\
&\vdots\\
s_M &= \frac{\partial \ln c}{\partial \ln p_M} = \beta_M + \delta_{M1}\ln p_1 + \delta_{M2}\ln p_2 + \cdots + \delta_{MM}\ln p_M.
\end{aligned}
\tag{10-37}
$$

The cost shares must sum to 1, which requires,

$$
\begin{aligned}
&\beta_1 + \beta_2 + \cdots + \beta_M = 1,\\
&\sum_{i=1}^{M} \delta_{ij} = 0 \quad \text{(column sums equal zero)},\\
&\sum_{j=1}^{M} \delta_{ij} = 0 \quad \text{(row sums equal zero)}.
\end{aligned}
\tag{10-38}
$$

We will also impose the (theoretical) symmetry restriction, $\delta_{ij} = \delta_{ji}$.

KT The system of **share equations** provides a seemingly unrelated regressions model that can be used to estimate the parameters of the model.[30] To make the model opera- KT
FN 30 tional, we must impose the restrictions in (10-38) and solve the problem of **singularity of the disturbance covariance matrix** of the share equations. The first is accomplished
KT by dividing the first $M-1$ prices by the Mth, thus eliminating the last term in each row and column of the parameter matrix. As in the Cobb–Douglas model, we obtain a nonsingular system by dropping the Mth share equation. We compute maximum likelihood estimates of the parameters to ensure **invariance** KT with respect to the choice of which share equation we drop. For the translog cost function, the elasticities of substitution are particularly simple to compute once the parameters have been estimated:

$$
\theta_{ij} = \frac{\delta_{ij} + s_i s_j}{s_i s_j}, \quad \theta_{ii} = \frac{\delta_{ii} + s_i(s_i - 1)}{s_i^2}. \tag{10-39}
$$

FN 31 These elasticities will differ at every data point. It is common to compute them at some central point such as the means of the data.[31]

10.3 *Example* **A Cost Function for U.S. Manufacturing**

A number of recent studies using the translog methodology have used a four-factor model, with capital K, labor L, energy E, and materials M, the factors of production. Among the first studies to employ this methodology was Berndt and Wood's (1975) estimation of a translog

[30] The cost function may be included, if desired, which will provide an estimate of β_0 but is otherwise inessential. Absent the assumption of constant returns to scale, however, the cost function will contain parameters of interest that do not appear in the share equations. As such, one would want to include it in the model. See Christensen and Greene (1976) for an application.

[31] They will also be highly nonlinear functions of the parameters and the data. A method of computing asymptotic standard errors for the estimated elasticities is presented in Anderson and Thursby (1986). Krinsky and Robb (1986, 1990) (see ~~Example 4.8~~ Section 15.3) proposed their method as an alternative approach to this computation.

TABLE 10.~~4~~ 3 Parameter Estimates (standard errors in parentheses)

β_K	0.05682	(0.00131)	δ_{KM}	−0.02169*	(0.00963)
β_L	0.25355	(0.001987)	δ_{LL}	0.07488	(0.00639)
β_E	0.04383	(0.00105)	δ_{LE}	−0.00321	(0.00275)
β_M	0.64580*	(0.00299)	δ_{LM}	−0.07169*	(0.00941)
δ_{KK}	0.02987	(0.00575)	δ_{EE}	0.02938	(0.00741)
δ_{KL}	0.0000221	(0.00367)	δ_{EM}	−0.01797*	(0.01075)
δ_{KE}	−0.00820	(0.00406)	δ_{MM}	0.11134*	(0.02239)

*Estimated indirectly using (10-38).

cost function for the U.S. manufacturing sector. The three factor shares used to estimate the model are

$$s_K = \beta_K + \delta_{KK}\ln\left(\frac{p_K}{p_M}\right) + \delta_{KL}\ln\left(\frac{p_L}{p_M}\right) + \delta_{KE}\ln\left(\frac{p_E}{p_M}\right),$$

$$s_L = \beta_L + \delta_{KL}\ln\left(\frac{p_K}{p_M}\right) + \delta_{LL}\ln\left(\frac{p_L}{p_M}\right) + \delta_{LE}\ln\left(\frac{p_E}{p_M}\right),$$

$$s_E = \beta_E + \delta_{KE}\ln\left(\frac{p_K}{p_M}\right) + \delta_{LE}\ln\left(\frac{p_L}{p_M}\right) + \delta_{EE}\ln\left(\frac{p_E}{p_M}\right).$$

F10.1

TB 10.3

Berndt and Wood's data are reproduced in Appendix Table F10.2. Constrained FGLS estimates of the parameters presented in Table 10.~~4~~ 3 were obtained by constructing the "pooled regression" in (10-19) with data matrices

$$\mathbf{y} = \begin{bmatrix} \mathbf{s_K} \\ \mathbf{s_L} \\ \mathbf{s_E} \end{bmatrix}, \tag{10-40}$$

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

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

FN 32

Estimates are then obtained using the two-step procedure in (10-7) and (10-9).[32] The full set of estimates are given in Table 10.~~5~~ 4. The parameters not estimated directly in (10-36) are computed using (10-38).

TB 10.4

FN 33

The implied estimates of the elasticities of substitution and demand for 1959 (the central year in the data) are derived in Table 10.~~5~~ 4 using the fitted cost shares and the estimated parameters in (10-39). The departure from the Cobb–Douglas model with unit elasticities is substantial. For example, the results suggest almost no substitutability between energy and labor and some complementarity between capital and energy.[33]

[32] These estimates do not match those reported by Berndt and Wood. They used an iterative estimator whereas ours is two steps FGLS. To purge their data of possible correlation with the disturbances, they first regressed the prices on 10 exogenous macroeconomic variables, such as U.S. population, government purchases of labor services, real exports of durable goods and U.S. tangible capital stock, and then based their analysis on the fitted values. The estimates given here are, in general quite close to those given by Berndt and Wood. For example, their estimates of the first five parameters are 0.0564, 0.2539, 0.0442, 0.6455, and 0.0254.

[33] Berndt and Wood's estimate of θ_{EL} for 1959 is 0.64.

10-32

TABLE 10.~~5~~ 4 Estimated Elasticities

	Capital	Labor	Energy	Materials
	Cost Shares for 1959			
Fitted shares	0.05646	0.27454	0.04424	0.62476
Actual shares	0.06185	0.27303	0.04563	0.61948
	Implied Elasticities of Substitution, 1959			
Capital	−7.34124			
Labor	1.0014	−1.64902		
Energy	−2.28422	0.73556	−6.59124	
Materials	0.34994	0.58205	0.38512	−0.19702
	0.38512		0.34994	−.31536
	Implied Own Price Elasticities			
	−0.41448	−0.45274	−0.29161	−0.19702

10.5 SUMMARY AND CONCLUSIONS

This chapter has surveyed use of the seemingly unrelated regressions model. The SUR model is an application of the generalized regression model introduced in Chapter 8. The advantage of the SUR formulation is the rich variety of behavioral models that fit into this framework. We began with estimation and inference with the SUR model, treating it essentially as a generalized regression. The major difference between this set of results and the single equation model in Chapter 8 is practical. While the SUR model is, in principle a single equation GR model with an elaborate covariance structure, special problems arise when we explicitly recognize its intrinsic nature as a set of equations linked by their disturbances. The major result for estimation at this step is the feasible GLS estimator. In spite of its apparent complexity, we can estimate the SUR model by a straightforward two-step GLS approach that is similar to the one we used for models with heteroscedasticity in Chapter 8. We also extended the SUR model to autocorrelation and heteroscedasticity. Once again, the multiple equation nature of the model complicates these applications. Maximum likelihood is an alternative method that is useful for systems of demand equations. This chapter examined a number of applications of the SUR model. Some panel data applications were presented in Section 10.3. Section 10.4 presented one of the most common recent applications of the seemingly unrelated regressions model, the estimation of demand systems. One of the signature features of this literature is the seamless transition from the theoretical models of optimization of consumers and producers to the sets of empirical demand equations derived from Roy's identity for consumers and Shephard's lemma for producers.

Key Terms and Concepts

- Autocorrelation
- Balanced panel
- Cobb-Douglas model
- Constant returns to scale
- Covariance structures model
- Demand system
- Feasible GLS
- Fixed effects
- Flexible functional form
- Generalized regression model
- Heteroscedasticity
- Homogeneity restriction
- Identical explanatory variables
- Identical regressors
- Invariance
- Kronecker product

10.6 SIMULTANEOUS EQUATIONS MODELS

chapter

There is a qualitative difference between the market equilibrium model suggested in the Introduction,

AU: Subs ok italic?

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

and the other examples considered thus far. The seemingly unrelated regression model,

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

derives from a set of regression equations that are connected through the disturbances. The regressors, $\mathbf{x}_{im}$ are exogenous and vary autonomously for reasons that are not explained within the model. Thus, the coefficients are directly interpretable as partial effects and can be estimated by least squares or other methods that are based on the conditional mean functions, $E[y_{im}|\mathbf{x}_{im}] = \mathbf{x}_{im}'\boldsymbol{\beta}$. In a model such as the equilibrium model above, the relationships are explicit and neither of the two market equations is a regression model. As a consequence, the partial equilibrium experiment of changing the price and inducing a change in the equilibrium quantity so as to elicit an estimate of the price elasticity of demand, α_2 (or supply elasticity, β_2) makes no sense. The model is of the joint determination of quantity and price. Price changes when the market equilibrium changes, but that is induced by changes in other factors, such as changes in incomes or other variables that affect the supply function. (See Figure 8.1 for a graphical treatment.) preceding

As we saw in Example 8.4, least squares regression of observed equilibrium quantities on price and the other factors will compute an ambiguous mixture of the supply and demand functions. The result follows from the endogeneity of Price in either equation. "Simultaneous equations models," arise in settings such as this one, in which the set of equations are interdependent by design. Simultaneous equations models will fit in the framework developed in Chapter 8, where we considered equations in which some of the right hand side variables are endogenous – that is, correlated with the disturbances. The substantive difference at this point is the source of the endogeneity. In our treatments in Chapter 8, endogeneity arose, for example in the models of omitted variables, measurement error, or endogenous treatment effects, essentially as an unintended deviation from the assumptions of the linear regression model. In the simultaneous equations framework, endogeneity is a fundamental part of the specification. This section will consider the issues of specification and estimation in systems of simultaneous equations. We begin in Section 10.6.1 with a development of a general framework for the analysis and a statement of some fundamental issues. Section 10.6.2 presents the simultaneous equations model as an extension of the seemingly unrelated regressions model in Section 10.2. The ultimate objective of the analysis will be to learn about the model coefficients. The issue of whether this is even possible is considered in Section 10.6.3, where we develop the issue of identification. Once the identification question is settled, methods of estimation and inference are presented in Section 10.6.4 and 10.6.5.

10.6.1 SYSTEMS OF EQUATIONS

preceding

Consider a simplified version of the equilibrium model above,

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

10-34

AU: Check numbering of footnotes.

not a new paragraph

These equations are **structural equations** in that they are derived from theory and each purports to describe a particular aspect of the economy.[2] Because the model is one of the joint determination of price and quantity, they are labeled **jointly dependent** or **endogenous** variables. Income, x, is assumed to be determined outside of the model, which makes it **exogenous.** The disturbances are added to the usual textbook description to obtain an **econometric model.** All three equations are needed to determine the equilibrium price and quantity, so the system is **interdependent.** Finally, because an equilibrium solution for price and quantity in terms of income and the disturbances is, indeed, implied (unless α_1 equals β_1), the system is said to be a **complete system of equations.** *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

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

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

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

10-41

(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 x and p. This result is characteristic of simultaneous-equations models. Because the endogenous variables are all correlated with the disturbances, the least squares estimators of the parameters of equations with endogenous variables on the right-hand side are inconsistent.[3]

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

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

[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.**

10-35

36 35

FN 36 KT KT

Since least squares is inconsistent, we might instead use an **instrumental variable estimator.**[35] The only variable in the system that is not correlated with the disturbances is x. Consider, then, the IV estimator, $\hat{\beta}_1 = \mathbf{q}'\mathbf{x}/\mathbf{p}'\mathbf{x}$. This estimator has

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

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

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

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

It would 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.~~ section.

KT

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

Example ~~13.1~~ 10-7 A Small Macroeconomic Model

Consider the model

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

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.~~

36 35 See Section ~~12.1~~ 8.3.

10-36

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. 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).

KT 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 $\mathbf{y}_t|\mathbf{x}_t$ 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 $s \geq 0$. Variables that are predetermined in a model can be treated, at least asymptotically, as if they were exogenous in the sense that consistent estimators can be derived when they appear as regressors. We will use this result in Chapter 21, when we derive the properties of regressions containing lagged values of the dependent variable. A related concept is KT **Granger (1969)-Sims (1977) causality.** Granger causality (a kind of statistical feedback) is absent when $f(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{t-1})$ equals $f(\mathbf{x}_t|\mathbf{x}_{t-1})$. 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.** KT

AU: KT "Granger $\frac{1}{N}$ Sims causality" is not in chap. list

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 $s \geq 0$. Variables that are predetermined in a model can be treated, at least asymptotically, as if they were exogenous in the sense that consistent estimators can be derived when they appear as regressors. We used this result in Section 12.8.2 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.**

~~13.2.2~~ 10.6.2 A GENERAL NOTATION FOR LINEAR SIMULTANEOUS EQUATIONS MODELS[~~36~~ 37]

FN 37

KT

FN 38

The **structural form** of the model is[~~37~~ 38]

$$
\begin{aligned}
\gamma_{11}y_{t1} + \gamma_{21}y_{t2} + \cdots + \gamma_{M1}y_{tM} + \beta_{11}x_{t1} + \cdots + \beta_{K1}x_{tK} &= \varepsilon_{t1},\\
\gamma_{12}y_{t1} + \gamma_{22}y_{t2} + \cdots + \gamma_{M2}y_{tM} + \beta_{12}x_{t1} + \cdots + \beta_{K2}x_{tK} &= \varepsilon_{t2},\\
&\vdots\\
\gamma_{1M}y_{t1} + \gamma_{2M}y_{t2} + \cdots + \gamma_{MM}y_{tM} + \beta_{1M}x_{t1} + \cdots + \beta_{KM}x_{tK} &= \varepsilon_{tM}.
\end{aligned}
\qquad 10\text{-}42\ (\cancel{13}/2)
$$

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

KT

KT

~~36~~ 37 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 (2004), and Wooldridge (2002a) provide further discussion.

~~37~~ 38 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.

10-38

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

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

Each column of the parameter matrices is the vector of coefficients in a particular equation, whereas each row applies to a specific endogenous variable.

The underlying theory will imply a number of restrictions on $\boldsymbol{\Gamma}$ 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 $\boldsymbol{\Gamma}$. 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 $\boldsymbol{\Gamma}$ and $\mathbf{B}$ will be completely known, and there will be no disturbance for that equation. Because not all variables appear in all equations, some of the parameters will be zero. The theory may also impose other types of restrictions on the parameter matrices.

If $\boldsymbol{\Gamma}$ 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.

10-39

KT 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,

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

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

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

KT

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

Example 13.3 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

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

10-40

Example 10.4 Structure and Reduced Form in 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 based on output, y_t, and one lagged value, an investment equation based on interest, r_t and the growth in output, and an equilibrium condition. The model determines the values of the three endogenous variables c_t, i_t, and y_t. This model is a **dynamic model.** In addition to the exogenous variables r_t and government spending, g_t, it contains two **predetermined variables,** c_{t-1} and y_{t-1}. These are obviously not exogenous, but with regard to the current values of the endogenous variables, they may be regarded as having already been determined. The deciding factor is whether or not they are uncorrelated with the current disturbances, which we might assume. The reduced form of this model is

KT KT

Ao: KT "predetermined variables" already KT in chap. Here also?

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

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

$$Ay_t = \alpha_0 + \beta_0 + \beta_1 r_t + g_t + \alpha_2 c_{t-1} - \beta_2 y_{t-1} + \varepsilon_{t1} + \varepsilon_{t2},$$

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

Denote $\mathbf{y}' = [c, i, y]$, $\mathbf{x}' = [1, r, g, c_{-1}, y_{-1}]$, and

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

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

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

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

10-41

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,

$$
\begin{aligned}
\mathbf{y}_t' &= [x_1 \quad x_2 \quad \cdots \quad x_K]_t \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1M} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2M} \\ & & \vdots & \\ \pi_{K1} & \pi_{K2} & \cdots & \pi_{KM} \end{bmatrix} + [\nu_1 \quad \cdots \quad \nu_M]_t \\
&= -\mathbf{x}_t'\mathbf{B}\boldsymbol{\Gamma}^{-1} + \boldsymbol{\varepsilon}_t'\boldsymbol{\Gamma}^{-1} \\
&= \mathbf{x}_t'\boldsymbol{\Pi} + \mathbf{v}_t'.
\end{aligned}
$$

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

Example 13.3 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

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

For the present, we assume that

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

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

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

This implies that

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

AQ: Term "reduced-form disturbances" already KT in chap. Here also?

10-42

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

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

An important assumption, comparable with the one made in Chapter 4 for the classical regression model, is

$$\operatorname{plim}(1/T)\mathbf{X}'\mathbf{X} = \mathbf{Q}, \quad \text{a finite positive definite matrix.} \qquad (\cancel{13\text{-}3}) \; 10\text{-}43$$

We also assume that

$$\operatorname{plim}(1/T)\mathbf{X}'\mathbf{E} = \mathbf{0}. \qquad (\cancel{13\text{-}4}) \; 10\text{-}44$$

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

$$\mathbf{Y} = \mathbf{X\Pi} + \mathbf{V}, \quad \text{where } \mathbf{V} = \mathbf{E\Gamma}^{-1}. \qquad (10\text{-}45)$$

Combining the earlier results, we have

$$\operatorname{plim} \frac{1}{T} \begin{bmatrix} \mathbf{Y}' \\ \mathbf{X}' \\ \mathbf{V}' \end{bmatrix} [\mathbf{Y} \quad \mathbf{X} \quad \mathbf{V}] = \begin{bmatrix} \mathbf{\Pi}'\mathbf{Q\Pi} + \mathbf{\Omega} & \mathbf{\Pi}'\mathbf{Q} & \mathbf{\Omega} \\ \mathbf{Q\Pi} & \mathbf{Q} & \mathbf{0}' \\ \mathbf{\Omega} & \mathbf{0} & \mathbf{\Omega} \end{bmatrix}.$$

[illegible]

13.3 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).

10-43

AU: Term "identification problem" is not in chap. list.

10.6.3 THE PROBLEM OF IDENTIFICATION

KT

Solving the **identification problem** logically precedes estimation. It is a crucial element of the model specification step. The issue is whether there is *any* way to obtain estimates of the parameters of the specified model. We have in hand a certain amount of information to use for inference about the 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. We have already encountered this problem in Chapter 4, where we examined the issue of *multicollinearity*. The "model,"

$$consumption = \beta_1 + \beta_2 WageIncome + \beta_3 NonWageIncome + \beta_4 TotalIncome + \varepsilon, \quad \textbf{(10-46)}$$

cannot be distinguished from the alternative model

$$consumption = \gamma_1 + \gamma_2 WageIncome + \gamma_3 NonWageIncome + \gamma_4 TotalIncome + w, \quad \textbf{(10-47)}$$

where $\gamma_1 = \beta_1$, $\gamma_2 = \beta_2 + a$, $\gamma_3 = \beta_3 + a$, $\gamma_4 = \beta_4 - a$ for some nonzero a, if the data consist only of consumption and the two income values (and their sum). However, if we know that if β_4 equals zero, then, as we saw in Chapter 4, γ_2 must equal β_2 and γ_3 must equal β_3. The additional information serves to rule out the alternative model. The notion of observational equivalence relates to what can be learned from the available information, which consists of the sample data and the restrictions that theory places on the equations of the model. In Chapter 8, where we examined the instrumental variable estimator, we defined identification in terms of sufficient moment equations. Indeed, Figure 8.1 is precisely an application of the principle of observational equivalence. The case of measurement error that we examined in Section 8.5 is likewise about identification. The sample regression coefficient, b, converges to a function of two underlying parameters, β and σ_u^2; plim $b = \beta/[1+\sigma_u^2/Q^{**}]$ where $Q^{**} = \text{plim}(\mathbf{x}^{*\prime}\mathbf{x}^{*}/n)$. With no further information about σ_u^2, we cannot infer β from the sample information, b and Q^{**} – by setting the differential, $db = 0$, you can see that there are different pairs of β and σ_u^2 that produce the same plim b.

A mathematical statement of the idea can be made in terms of the likelihood function, which embodies the sample information. At this point, it helps to drop the statistical distinction between "**y**" and "**x**" and consider, in generic terms, the joint probability distribution for the observed data, $p(\mathbf{Y},\mathbf{X}|\boldsymbol{\theta})$, given the model parameters. Two model structures are observationally equivalent if

$$p(\mathbf{Y},\mathbf{X}|\boldsymbol{\theta}_1) = p(\mathbf{Y},\mathbf{X}|\boldsymbol{\theta}_2) \text{ for } \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \text{ for all realizations of } (\mathbf{Y},\mathbf{X}).$$

FN 39

A structure is said to be *unidentified* if it is observationally equivalent to another structure.[38] 39 (For our preceding consumption example, as will usually be the case when a model is unidentified, there are an infinite number of structures that are all equivalent to (10-46), one for each nonzero value of a in (10-47).

39

[38] See Hsiao (1983) for a survey of this issue.

10-44

The general simultaneous equations model we have specified in (10-42) is not identified. We have implicitly assumed that the marginal distribution of **X** can be separated from the conditional distribution of **Y**|**X**. We can write the model as

minus minus

$$p(\mathbf{Y},\mathbf{X}|\boldsymbol{\Gamma},\mathbf{B},\boldsymbol{\Sigma},\boldsymbol{\Theta}) = p(\mathbf{Y}|\mathbf{X},\boldsymbol{\Pi},\boldsymbol{\Omega})p(\mathbf{X}|\boldsymbol{\Theta}) \text{ with } \boldsymbol{\Pi} = -\mathbf{B}\boldsymbol{\Gamma}^{-1} \text{ and } \boldsymbol{\Omega} = (\boldsymbol{\Gamma}')^{-1}\boldsymbol{\Sigma}(\boldsymbol{\Gamma})^{-1}.$$

We assume that $\boldsymbol{\Theta}$ and $(\boldsymbol{\Gamma},\mathbf{B},\boldsymbol{\Sigma})$ have no elements in common. But, let **F** be any nonsingular $M\times M$ matrix and define $\mathbf{B}_2 = \mathbf{FB}$ and $\boldsymbol{\Gamma}_2 = \mathbf{F}\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}_2 = \mathbf{F}'\boldsymbol{\Sigma}\mathbf{F}$ (i.e., we just multiply the whole model by **F**). If **F** is not equal to an identity matrix. Then $\mathbf{B}_2$, $\boldsymbol{\Gamma}_2$ and $\boldsymbol{\Sigma}_2$ are a different **B**, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ that are consistent with the same data, that is, with the same (**Y**,**X**) which imply ($\boldsymbol{\Pi}$ and $\boldsymbol{\Omega}$). This follows because $\boldsymbol{\Pi}_2 = -\mathbf{B}_2^{-1}\boldsymbol{\Gamma}_2 = -\mathbf{B}^{-1}\boldsymbol{\Gamma} = \boldsymbol{\Pi}$ and likewise for $\boldsymbol{\Omega}_2$. To see how this will proceed from here, consider that in each equation, there is one "dependent variable," that is a variable whose coefficient equals one. Therefore, one specific element of $\boldsymbol{\Gamma}$ in every equation (column) equals one. That rules out any matrix **F** which does not leave a one in that position in $\boldsymbol{\Gamma}_2$. Likewise, in the market equilibrium case in Section 10.6.1, the coefficient on x in the supply equation is zero. That means there is an element in one of the columns of **B** that equals zero. Any **F** that does not preserve that zero restriction is invalid. Thus, certain restrictions that theory imposes on the model rule out some of the alternative models. With enough restrictions, the only valid **F** matrix will be **F** = **I**, and the model becomes identified.

minus

The structural model consists of the equation system

minus

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

Each column in $\boldsymbol{\Gamma}$ and **B** are the parameters of a specific equation in the system. The sample information consists of, at the first instance the data, (**Y**,**X**), and other nonsample information in the form of restrictions on parameter matrices, such as the normalizations noted in the preceding example. The sample data provide sample moments, $\mathbf{X}'\mathbf{X}/n$, $\mathbf{X}'\mathbf{Y}/n$ and $\mathbf{Y}'\mathbf{Y}/n$. For purposes of identification, which is independent of issues of sample size, suppose we could observe as large a sample as desired. Then, we could observe [from (10-45)]

$$\begin{aligned}
&\text{plim}(1/n)\mathbf{X}'\mathbf{X} = \mathbf{Q},\\
&\text{plim}(1/n)\mathbf{X}'\mathbf{Y} = \text{plim}(1/n)\mathbf{X}'(\mathbf{X}\boldsymbol{\Pi} + \mathbf{V}) = \mathbf{Q}\boldsymbol{\Pi},\\
&\text{plim}(1/n)\mathbf{Y}'\mathbf{Y} = \text{plim}(1/n)(\mathbf{X}\boldsymbol{\Pi} + \mathbf{V})'(\mathbf{X}\boldsymbol{\Pi} + \mathbf{V}) = \boldsymbol{\Pi}'\mathbf{Q}\boldsymbol{\Pi} + \boldsymbol{\Omega}.
\end{aligned}$$

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

$$\boldsymbol{\Pi} = [\text{plim}(1/n)\mathbf{X}'\mathbf{Y}]^{-1}[\text{plim}(1/n)\mathbf{X}'\mathbf{Y}]$$

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

$$\boldsymbol{\Omega} = [\text{plim}(1/n)\mathbf{Y}'\mathbf{Y}] - [\text{plim}(1/n)\mathbf{Y}'\mathbf{X}]\,[\text{plim}(1/n)\mathbf{X}'\mathbf{X}]^{-1}[\text{plim}(1/n)\mathbf{X}'\mathbf{Y}].$$

This result should be recognized as the matrix of least squares residual variances and covariances. Therefore,

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

10-45

The information in hand, therefore, consists of **Π**, **Ω**, and whatever other nonsample information we have about the structure.[38]

FN 40

Thus, **Π** and **Ω** are "observable." ~~What interests us is **Γ**,**B**,**Σ**.~~ The ultimate question is whether we can deduce **Γ**,**B**,**Σ** from **Π**,**Ω**. A simple counting exercise immediately reveals that the answer is no — there are M^2 parameters **Γ**, $M(M+1)/2$ in **Σ** and KM in **B** to be deduced. The sample data contain KM elements in **Π** and $M(M+1)/2$ elements in **Ω**. By simply counting equations and unknowns, we find that our data are insufficient by M^2 pieces of information. We have (in principle) used the sample information already, so these M^2 additional restrictions are going to be provided by the theory of the model. A small example will help to fix ideas.

Example ~~23~~10.5 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,

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

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

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

40

[38] We have not necessarily shown that this is *all* the information in the sample. In general, we observe the conditional distribution $f(\mathbf{y}_i|\mathbf{x}_i)$, which constitutes the likelihood for the reduced form. With normally distributed disturbances, this distribution is a function of only **Π** and **Ω**. 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 Reirsøl (1950).

note accent

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

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

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.

KT

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

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

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

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

AU: Term "rank and order conditions" not in chap. list

The conclusion is that some equation systems are identified and others are not. The formal mathematical conditions under which an equation system is identified turns on some intricate results known as the **rank and order conditions**. KT

The *order condition* is a simple counting rule. In the equation system context, the order condition is that the number of exogenous variables that appear elsewhere in the equation system must be at least as large as the number of endogenous variables in the equation. We used this rule when we constructed the IV estimator in Chapter 8. In that setting, we required our model to be at least "identified" by requiring that the number of instrumental variables not contained in **X** be at least as large as the number of endogenous variables. The correspondence of that single equation application with the condition defined here is that the rest of the equation system is, essentially, the rest of the world (i.e., the source of the instrumental variables).[39] A simple sufficient order condition for an equation system is that each equation must contain "its own" exogenous variable that does not appear elsewhere in the system.

FN 41

41

[39] This invokes the perennial question (encountered repeatedly in the applications in Chapter 8), "where do the instruments come from?" See Section 8.8 for discussion.

41

minus

The order condition is necessary for identification; the *rank condition* is sufficient. The The equation system in (10-42) in structural form is $\mathbf{y}'\mathbf{\Gamma} = -\mathbf{x}'\mathbf{B} + \boldsymbol{\varepsilon}'$. The reduced form is $\mathbf{y}' = \mathbf{x}'(-\mathbf{B}\mathbf{\Gamma}^{-1}) + \boldsymbol{\varepsilon}'\mathbf{\Gamma}^{-1} = \mathbf{x}'\mathbf{\Pi} + \mathbf{v}'$. The way we are going to deduce the parameters in $(\mathbf{\Gamma},\mathbf{B},\mathbf{\Sigma})$ is from the reduced form parameters $(\mathbf{\Pi},\mathbf{\Omega})$. For a particular equation, say the *j*th, the solution is contained in $\mathbf{\Pi\Gamma} = -\mathbf{B}$, or for a particular equation, $\mathbf{\Pi\Gamma}_j = -\mathbf{B}_j$ where $\mathbf{\Gamma}_j$ contains all the coefficients in the *j*th equation that multiply endogenous variables. One of these coefficients will equal one, usually some will equal zero, and the remainder are the nonzero coefficients on endogenous variables in the equation, $\mathbf{Y}_j$ [these are denoted $\boldsymbol{\gamma}_j$ in (10-48) following]. Likewise, $\mathbf{B}_j$ contains the coefficients in equation *j* on all exogenous variables in the model – some of these will be zero and the remainder will multiply variables in $\mathbf{X}_j$, the exogenous variables that appear in this equation [these are denoted $\boldsymbol{\beta}_j$ in (10-48) following]. The empirical counterpart will be

minus minus minus minus

$$[\text{plim}(1/n)\mathbf{X}'\mathbf{X}]^{-1}[\text{plim}(1/n)\mathbf{X}'\mathbf{Y}_j]\,\mathbf{\Gamma}_j - \mathbf{B}_j = \mathbf{0}.$$

The rank condition ~~insures~~ ensures that there is a unique solution to this set of equations. In practical terms, the rank condition is difficult to establish in large equation systems. Practitioners typically take it as a given. In small systems, such as the 2 or 3 equation systems that dominate contemporary research, it is trivial. We have already used the rank condition in Chapter 8 where it played a role in the "relevance" condition for instrumental variable estimation. In particular, note after after the statement of the assumptions for instrumental variable estimation, we assumed $\text{plim}(1/n)\mathbf{Z}'\mathbf{X}$ is a matrix with rank K. (This condition is often labeled the "rank condition" in contemporary applications. It not identical, but it is sufficient for the condition mentioned here.)

To add all this up, it is instructive to return to the order condition. We are trying to solve a set of moment equations based on the relationship between the structural parameters and the reduced form. The sample information provides $KM + M(M+1)/2$ items in $\mathbf{\Pi}$ and $\mathbf{\Omega}$. We require M^2 additional restrictions, imposed by the theory behind the model. The restrictions come in the form of normalizations, most commonly exclusion restrictions, and other relationships among the parameters, such as linear relationships, or specific values attached to coefficients.

The question of identification is a theoretical exercise. It arises in all econometric settings in which the parameters of a model are to be deduced from the combination of sample information and nonsample (theoretical) information. The crucial issue in each of these cases is our ability (or lack of) to deduce the values of structural parameters uniquely from sample information in terms of sample moments coupled with nonsample information, mainly restrictions on parameter values. The issue of identification is the subject of a lengthy literature including Working (1927) (which has been adapted to produce Figure 8.1), Gabrielsen (1978), Amemiya (1985), Bekker and Wansbeek (2001), and continuing through the contemporary discussion of natural experiments (Section 8.8 and Angrist and Pischke (2010), with commentary).

10-48

Single Equation

10.6.4 Estimation and Inference

For purposes of estimation and inference, we write the specification of the simultaneous equations model in the form that the researcher would typically formulate it;

$$\begin{aligned} \mathbf{y}_j &= \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{Y}_j\boldsymbol{\gamma}_j + \boldsymbol{\varepsilon}_j \\ &= \mathbf{Z}_j\boldsymbol{\delta}_j + \boldsymbol{\varepsilon}_j \end{aligned} \tag{10-48}$$

where $\mathbf{y}_j$ is the "dependent variable" in the equation, $\mathbf{X}_j$ is the set of exogenous variables that appear in the *j*th equation – note that this is not all the variables in the model – and $\mathbf{Z}_j = (\mathbf{X}_j,\mathbf{Y}_j)$. The full set of exogenous variables in the model, including $\mathbf{X}_j$ and variables that appear elsewhere in the model (including a constant term if any equation includes one) is denoted $\mathbf{X}$. For example, in the supply/demand model in Example 10.5, the full set of exogenous variables is $\mathbf{X} = (\mathbf{1},\mathbf{x},\mathbf{z})$, while for the demand equation, $\mathbf{X}_{Demand} = (\mathbf{1},\mathbf{x})$ and $\mathbf{X}_{Supply} = (\mathbf{1},\mathbf{z})$. Finally, $\mathbf{Y}_j$ is the endogenous variables that appear on the right hand side of the *j*th equation. Once again, this is likely to be a subset of the endogenous variables in the full model. In Example 10.5, $\mathbf{Y}_j$ = **(price)** in both cases.

Au: Are italic subs OK?

Au: Is bold "(price)" OK?

(KT) There are two approaches to estimation and inference for simultaneous equations models. **Limited information estimators** are constructed for each equation individually. The approach is analogous to estimation of the seemingly unrelated regressions model in Section 10.2 by least squares, one equation at a time. **Full information estimators** are used to estimate all equations simultaneously. The counterpart for the seemingly unrelated regressions model is the feasible generalized least squares estimator discussed in Section 10.2.3. The major difference to be accommodated at this point is the endogeneity of $\mathbf{Y}_j$ in (10-48).

The equation system in (10-48) is precisely the model developed in Chapter 8. Least squares will generally be unsuitable as it is inconsistent due to the correlation between $\mathbf{Y}_j$ and $\boldsymbol{\varepsilon}_j$. The usual approach will be two stage least squares as developed in Section 8.3.2 to 8.3.4. The only difference between the case considered here and that in Chapter 8 is the source of the instrumental variables. In our general model in Chapter 8, the source of the instruments remained somewhat ambiguous; the overall rule was "outside the model." In this setting, the instruments come from elsewhere in the model – that is, "not in the jth equation." Thus, for estimating the linear simultaneous equations model, the most common estimator is

$$\begin{aligned} \hat{\boldsymbol{\delta}}_{j,2SLS} &= [\hat{\mathbf{Z}}_j'\hat{\mathbf{Z}}_j]^{-1}\hat{\mathbf{Z}}_j'\mathbf{y}_j \\ &= [(\mathbf{Z}_j'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}_j)]^{-1}(\mathbf{Z}_j'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_j, \end{aligned} \qquad (10\text{-}49)$$

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

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

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

$$\hat{\sigma}_{jj} = \frac{(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_j)'(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_j)}{T} \qquad (10\text{-}50)$$

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

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

Two stage least squares is used nearly universally in estimation of simultaneous equation models – for precisely the reasons outlined in Chapter 8. However, some applications (and some theoretical treatments) have suggested that the **limited information maximum likelihood (LIML) estimator** based on the normal distribution may have better properties. The technique has also found recent use in the analysis of weak instruments that we consider in Section 10.6.5. A full (lengthy) derivation of the log-likelihood is provided in Davidson and MacKinnon (2004). We will proceed to the practical aspects of this estimator and refer the reader to this source 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).] One significant virtue of LIML is its invariance to the normalization of the equation. Consider an example in a system of equations,

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

An equivalent equation would be

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

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 earlier. A second virtue is LIML's better performance in the presence of weak instruments.

10-50

FN 42

KT

The LIML, or **least variance ratio** estimator, can be computed as follows.~~[18]~~ 40 ~~42~~ Let

$$\mathbf{W}_j^0 = \mathbf{E}_j^{0\prime}\mathbf{E}_j^0,$$

~~(13-22)~~ 10-51

where

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

and

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

10-52 ~~(13-23)~~

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

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

10-53 ~~(13-24)~~

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

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

10-54 ~~(13-25)~~

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,

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

10-55 ~~(13-26)~~

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

40 ~~42~~ ~~[18]~~The least variance ratio estimator is derived in Johnston (1984). The LIML estimator was derived by Anderson and Rubin (1949, 1950). The LIML estimator has, since its derivation by Anderson and Rubin in 1949 and 1950, been of largely theoretical interest only. The much simpler and equally efficient two-stage least squares estimator has stood as the estimator of choice. But LIML and the A-R specification test have been rediscovered and reinvigorated with their use in the analysis of weak instruments. See Hahn and Hausman (2002, 2003) and Sections ~~12.9~~ and ~~13.5.5~~.

8.7 10.6.6

10-51

estimator.[41] The implication is that with normally distributed disturbances, 2SLS is fully efficient.

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

$$\hat{\boldsymbol{\delta}}_{j,k} = \begin{pmatrix} \hat{\boldsymbol{\gamma}}_{j,k} \\ \hat{\boldsymbol{\beta}}_{j,k} \end{pmatrix} = \begin{bmatrix} \mathbf{Y}_j'\mathbf{Y}_j - k\mathbf{V}_j'\mathbf{V}_j & \mathbf{Y}_j'\mathbf{X}_j \\ \mathbf{X}_j'\mathbf{Y}_j & \mathbf{X}_j'\mathbf{X}_j \end{bmatrix} \begin{bmatrix} \mathbf{Y}_j'\mathbf{y}_j - k\mathbf{V}_j'\mathbf{v}_j \\ \mathbf{X}_j'\mathbf{y}_j \end{bmatrix}, \tag{10-56}$$

where $\mathbf{V}_j$ and $\mathbf{v}_j$ are the reduced form disturbances in (10-45). The feasible estimator is computed using the residuals from the OLS regressions of $\mathbf{Y}_j$ and $\mathbf{y}_j$ on $\mathbf{X}$ (not $\mathbf{X}_j$). 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 (10-55).] There have been many other k-class estimators derived; Davidson and MacKinnon (2004, pp. 537–538 and 548–549) 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 estimators are the same, but the finite-sample properties are possibly very different. Davidson and MacKinnon (2004, pp. 537–538 and 548–549) 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.

[41] 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).

10-52

As noted earlier, analysis of problems of estimation and inference in structural equation models is one of the bedrock platforms of research in econometrics. The analysis sketched here continues that line of inquiry.

13.5.6 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.[17] A case that is broad enough to include many practical applications is the one analyzed by Kelejian (1971),

$$\mathbf{y}_j = \gamma_{1j}\mathbf{f}_{1j}(\mathbf{y}, \mathbf{x}) + \gamma_{2j}\mathbf{f}_{2j}(\mathbf{y}, \mathbf{x}) + \cdots + \mathbf{X}_j\boldsymbol{\beta}_j + \varepsilon_j,^{18}$$

which is an extension of (6-4). 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, because 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 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.

10.6.5 ~~13.6~~ SYSTEM METHODS OF ESTIMATION

We may formulate the full system of equations as

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \vdots \\ \boldsymbol{\delta}_M \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix} \qquad (10\text{-}57)$$

or

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

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

[18] 2SLS for models that are nonlinear in the parameters is discussed in Chapter 15 in connection with GMM estimators.

where

$$E[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}, \quad \text{and} \quad E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \mathbf{X}] = \overline{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} \otimes \mathbf{I}. \qquad \text{10-58 } \cancel{(13\text{-}28)}$$

[See (10-~~3~~6).] 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 Chapter 10~~ 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 10~~ Section 10.2.1, we use a generalized least squares approach. Thus, assuming that the matrix of instrumental variables, $\overline{\mathbf{W}}$ satisfies the requirements for an IV estimator, a consistent though inefficient estimator would be

$$\hat{\boldsymbol{\delta}}_{\text{IV}} = [\overline{\mathbf{W}}'\mathbf{Z}]^{-1}\overline{\mathbf{W}}'\mathbf{y}. \qquad \text{10-59 } \cancel{(13\text{-}29)}$$

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

$$\hat{\boldsymbol{\delta}}_{\text{IV,GLS}} = [\overline{\mathbf{W}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Z}]^{-1}\overline{\mathbf{W}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{y}, \qquad \text{10-60 } \cancel{(13\text{-}30)}$$

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

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

Three IV techniques are generally used for joint estimation of the e[illegible] system of equations: three-stage least squares, GMM, and full information maximum likelihood. We will consider three-stage least squares here. GMM and FIML are discussed in Chapters 13 and 14, respectively.

~~13.6.1 THREE-STAGE LEAST SQUARES~~

new paragraph

Consider the IV estimator formed from

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

The IV estimator,

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

is simply equation-by-equation 2SLS. We have already established the consistency of 2SLS. By analogy to the seemingly unrelated regressions model of ~~Chapter 10~~ Section 10.2, however, we would expect this estimator to be less efficient than a GLS estimator. A natural candidate would be

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

10-54

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

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

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

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

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

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

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

$$\hat{\boldsymbol{\delta}}_{3SLS}=[\hat{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I})\hat{\mathbf{Z}}]^{-1}\hat{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I})\mathbf{y}. \qquad \text{10-61 } (\cancel{13\text{-}31})$$

The appropriate asymptotic covariance matrix for the estimator is

$$\text{Asy. Var}[\hat{\boldsymbol{\delta}}_{3SLS}]=[\overline{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I})\overline{\mathbf{Z}}]^{-1}, \qquad \text{10-62 } (\cancel{13\text{-}32})$$

where $\overline{\mathbf{Z}}=\text{diag}[\mathbf{X}\boldsymbol{\Pi}_j,\mathbf{X}_j]$. This matrix would be estimated with the bracketed inverse matrix in (~~13-31~~ 10-61).

Using sample data, we find that $\overline{\mathbf{Z}}$ may be estimated with $\hat{\mathbf{Z}}$. The remaining difficulty is to obtain an estimate of $\boldsymbol{\Sigma}$. In estimation of the 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:

KT

AU: KT "estimator" is not in chap. list

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

$$\hat{\sigma}_{ij}=\frac{(\mathbf{y}_i-\mathbf{Z}_i\hat{\boldsymbol{\delta}}_i)'(\mathbf{y}_j-\mathbf{Z}_j\hat{\boldsymbol{\delta}}_j)}{T}. \qquad \text{10-63 } (\cancel{13\text{-}33})$$

3. Compute the GLS estimator according to (~~13-31~~ 10-61) and an estimate of the asymptotic covariance matrix according to (~~13-32~~ 10-62) 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.[~~19~~ 42 44]

FN 44

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

44 42 ~~19~~ A Jacobian term needed to maximize the log-likelihood is not treated by the 3SLS estimator. See Dhrymes (1973).

10-55

43 45

FN 45

be shown that among all IV estimators that use only the sample information embodied in the system, 3SLS is asymptotically efficient.[20] 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.

13.6.2 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. (Like the LIML estimator in Section 13.5.4, the FIML estimator for linear simultaneous equations models stands somewhat apart from the other maximum likelihood applications developed in Chapter 16. The practical interest in the estimator is rather limited because the 3SLS estimator is equally efficient, much easier to compute, and does not impose a normality assumption. On the other hand, like the LIML estimator, the FIML estimator presents a useful theoretical benchmark. As such, it is more useful to present it here, while the background theory for the ML methodology can be found in Chapter 16.)

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

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

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, $E[\mathbf{v}_t\mathbf{v}_t' \mid \mathbf{X}] = \boldsymbol{\Omega}$. The log-likelihood for this model is precisely that of the seemingly unrelated regressions model of Chapter 10. (See Section 16.9.3.b.) For the moment, we can ignore the relationship between the structural and reduced-form parameters. The log-likelihood function is

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

where

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

and

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

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

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

43 45

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