9 Simultaneous Equations Models

9.1 The Scope of Simultaneous Equations Models

The emphasis in this chapter is on situations where two or more variables are jointly determined by a system of equations. Nevertheless, the population model, the identification analysis, and the estimation methods apply to a much broader range of problems. In Chapter 8, we saw that the omitted variables problem described in Example 8.2 has the same statistical structure as the true simultaneous equations model in Example 8.1. In fact, any or all of simultaneity, omitted variables, and measurement error can be present in a system of equations. Because the omitted variable and measurement error problems are conceptually easier—and it was for this reason that we discussed them in single-equation contexts in Chapters 4 and 5—our examples and discussion in this chapter are geared mostly toward true **simultaneous equations models (SEMs)**.

For effective application of true SEMs, we must understand the kinds of situations suitable for SEM analysis. The labor supply and wage offer example, Example 8.1, is a legitimate SEM application. The labor supply function describes individual behavior, and it is derivable from basic economic principles of individual utility maximization. Holding other factors fixed, the labor supply function gives the hours of labor supply at *any* potential wage facing the individual. The wage offer function describes firm behavior, and, like the labor supply function, the wage offer function is self-contained.

When an equation in an SEM has economic meaning in isolation from the other equations in the system, we say that the equation is **autonomous**. One way to think about autonomy is in terms of counterfactual reasoning, as in Example 8.1. If we know the parameters of the labor supply function, then, for any individual, we can find labor hours given any value of the potential wage (and values of the other observed and unobserved factors affecting labor supply). In other words, we could, in principle, trace out the individual labor supply function for given levels of the other observed and unobserved variables.

Causality is closely tied to the autonomy requirement. An equation in an SEM should represent a causal relationship; therefore, we should be interested in varying each of the explanatory variables—including any that are endogenous—while holding *all* the others fixed. Put another way, each equation in an SEM should represent *some* underlying conditional expectation that has a causal structure. What complicates matters is that the conditional expectations are in terms of counterfactual variables. In the labor supply example, *if* we could run a controlled experiment, where we exogenously vary the wage offer across individuals, then the labor supply function could be estimated without ever considering the wage offer function. In fact, in the absence of omitted variables or measurement error, ordinary least squares would be an appropriate estimation method.

Generally, supply and demand examples satisfy the autonomy requirement, regardless of the level of aggregation (individual, household, firm, city, and so on), and simultaneous equations systems were originally developed for such applications. [See, for example, Haavelmo (1943) and Kiefer's (1989) interview of Arthur S. Goldberger.] Unfortunately, many recent applications of simultaneous equations methods fail the autonomy requirement; as a result, it is difficult to interpret what has actually been estimated. Examples that fail the autonomy requirement often have the same feature: the endogenous variables in the system are all choice variables of the *same* economic unit.

As an example, consider an individual's choice of weekly hours spent in legal market activities and hours spent in criminal behavior. An economic model of crime can be derived from utility maximization; for simplicity, suppose the choice is only between hours working legally (*work*) and hours involved in crime (*crime*). The factors assumed to be exogenous to the individual's choice are things like wage in legal activities, other income sources, probability of arrest, expected punishment, and so on. The utility function can depend on education, work experience, gender, race, and other demographic variables.

Two structural equations fall out of the individual's optimization problem: one has *work* as a function of the exogenous factors, demographics, and unobservables; the other has *crime* as a function of these same factors. Of course, it is always possible that factors treated as exogenous by the individual cannot be treated as exogenous by the econometrician: unobservables that affect the choice of *work* and *crime* could be correlated with the observable factors. But this possibility is an omitted variables problem. (Measurement error could also be an important issue in this example.) Whether or not omitted variables or measurement error are problems, each equation has a causal interpretation.

In the crime example, and many similar examples, it may be tempting to stop before completely solving the model—or to circumvent economic theory altogether—and specify a simultaneous equations system consisting of two equations. The first equation would describe *work* in terms of *crime*, while the second would have *crime* as a function of *work* (with other factors appearing in both equations). While it is often possible to write the first-order conditions for an optimization problem in this way, these equations are *not* the structural equations of interest. Neither equation can stand on its own, and neither has a causal interpretation. For example, what would it mean to study the effect of changing the market wage on hours spent in criminal activity, holding hours spent in legal employment fixed? An individual will generally adjust the time spent in both activities to a change in the market wage.

Often it *is* useful to determine how one endogenous choice variable trades off against another, but in such cases the goal is not—and should not be—to infer causality. For example, Biddle and Hamermesh (1990) present OLS regressions of minutes spent per week sleeping on minutes per week working (controlling for education, age, and other demographic and health factors). Biddle and Hamermesh recognize that there is nothing "structural" about such an analysis. (In fact, the choice of the dependent variable is largely arbitrary.) Biddle and Hamermesh (1990) *do* derive a structural model of the demand for sleep (along with a labor supply function) where a key explanatory variable is the wage offer. The demand for sleep has a causal interpretation, and it does *not* include labor supply on the right-hand side.

Why are SEM applications that do not satisfy the autonomy requirement so prevalent in applied work? One possibility is that there appears to be a general misperception that "structural" and "simultaneous" are synonymous. However, we already know that structural models need not be systems of simultaneous equations. And, as the crime/work example shows, a simultaneous system is not necessarily structural.

9.2 Identification in a Linear System

9.2.1 Exclusion Restrictions and Reduced Forms

Write a system of linear simultaneous equations for the population as

$$\begin{aligned} y_1 &= \mathbf{y}_{(1)}\gamma_{(1)} + \mathbf{z}_{(1)}\boldsymbol{\delta}_{(1)} + u_1 \\ &\vdots \\ y_G &= \mathbf{y}_{(G)}\gamma_{(G)} + \mathbf{z}_{(G)}\boldsymbol{\delta}_{(G)} + u_G \end{aligned} \tag{9.1}$$

where $\mathbf{y}_{(h)}$ is $1 \times G_h$, $\gamma_{(h)}$ is $G_h \times 1$, $\mathbf{z}_{(h)}$ is $1 \times M_h$, and $\boldsymbol{\delta}_{(h)}$ is $M_h \times 1$, $h = 1, 2, \ldots, G$. These are **structural equations** for the **endogenous variables** $y_1, y_2, \ldots, y_G$. We will assume that, if the system (9.1) represents a true simultaneous equations model, then equilibrium conditions have been imposed. Hopefully, each equation is autonomous, but, of course, they do not need to be for the statistical analysis.

The vector $\mathbf{y}_{(h)}$ denotes endogenous variables that appear on the right-hand side of the hth structural equation. By convention, $\mathbf{y}_{(h)}$ can contain any of the endogenous variables $y_1, y_2, \ldots, y_G$ *except* for y_h. The variables in $\mathbf{z}_{(h)}$ are the **exogenous variables** appearing in equation h. Usually there is some overlap in the exogenous variables across different equations; for example, except in special circumstances each $\mathbf{z}_{(h)}$ would contain unity to allow for nonzero intercepts. The restrictions imposed in system (9.1) are called **exclusion restrictions** because certain endogenous and exogenous variables are excluded from some equations.

The $1 \times M$ vector of all exogenous variables $\mathbf{z}$ is assumed to satisfy

$$\mathrm{E}(\mathbf{z}'u_g) = \mathbf{0}, \qquad g = 1, 2, \ldots, G \tag{9.2}$$

When all of the equations in system (9.1) are truly structural, we are usually willing to assume

$$\mathrm{E}(u_g \mid \mathbf{z}) = 0, \qquad g = 1, 2, \ldots, G \tag{9.3}$$

However, we know from Chapters 5 and 8 that assumption (9.2) is sufficient for consistent estimation. Sometimes, especially in omitted variables and measurement error applications, one or more of the equations in system (9.1) will simply represent a linear projection onto exogenous variables, as in Example 8.2. It is for this reason that we use assumption (9.2) for most of our identification and estimation analysis. We assume throughout that $\mathrm{E}(\mathbf{z}'\mathbf{z})$ is nonsingular, so that there are no exact linear dependencies among the exogenous variables in the population.

Assumption (9.2) implies that the exogenous variables appearing anywhere in the system are orthogonal to *all* the structural errors. If some elements in, say, $\mathbf{z}_{(1)}$, do not appear in the second equation, then we are explicitly assuming that they do not enter the structural equation for y_2. If there are no reasonable exclusion restrictions in an SEM, it may be that the system fails the autonomy requirement.

Generally, in the system (9.1), the error u_g in equation g will be correlated with $\mathbf{y}_{(g)}$ (we show this correlation explicitly later), and so OLS and GLS will be inconsistent. Nevertheless, under certain identification assumptions, we can estimate this system using the instrumental variables procedures covered in Chapter 8.

In addition to the exclusion restrictions in system (9.1), another possible source of identifying information is on the $G \times G$ variance matrix $\boldsymbol{\Sigma} \equiv \mathrm{Var}(\mathbf{u})$. For now, $\boldsymbol{\Sigma}$ is unrestricted and therefore contains no identifying information.

To motivate the general analysis, consider specific labor supply and demand functions for some population:

$$h^s(\omega) = \gamma_1 \log(\omega) + \mathbf{z}_{(1)}\boldsymbol{\delta}_{(1)} + u_1$$

$$h^d(\omega) = \gamma_2 \log(\omega) + \mathbf{z}_{(2)}\boldsymbol{\delta}_{(2)} + u_2$$

where ω is the dummy argument in the labor supply and labor demand functions. We assume that observed hours, h, and observed wage, w, equate supply and demand:

$h = h^s(w) = h^d(w)$

The variables in $\mathbf{z}_{(1)}$ shift the labor supply curve, and $\mathbf{z}_{(2)}$ contains labor demand shifters. By defining $y_1 = h$ and $y_2 = \log(w)$ we can write the equations in equilibrium as a linear simultaneous equations model:

$$y_1 = \gamma_1 y_2 + \mathbf{z}_{(1)}\boldsymbol{\delta}_{(1)} + u_1 \tag{9.4}$$

$$y_1 = \gamma_2 y_2 + \mathbf{z}_{(2)}\boldsymbol{\delta}_{(2)} + u_2 \tag{9.5}$$

Nothing about the general system (9.1) rules out having the same variable on the left-hand side of more than one equation.

What is needed to identify the parameters in, say, the supply curve? Intuitively, since we observe only the equilibrium quantities of hours and wages, we cannot distinguish the supply function from the demand function if $\mathbf{z}_{(1)}$ and $\mathbf{z}_{(2)}$ contain exactly the same elements. If, however, $\mathbf{z}_{(2)}$ contains an element *not* in $\mathbf{z}_{(1)}$—that is, if there is some factor that exogenously shifts the demand curve but not the supply curve—then we can hope to estimate the parameters of the supply curve. To identify the demand curve, we need at least one element in $\mathbf{z}_{(1)}$ that is not also in $\mathbf{z}_{(2)}$.

To formally study identification, assume that $\gamma_1 \neq \gamma_2$; this assumption just means that the supply and demand curves have different slopes. Subtracting equation (9.5) from equation (9.4), dividing by $\gamma_2 - \gamma_1$, and rearranging gives

$$y_2 = \mathbf{z}_{(1)}\boldsymbol{\pi}_{21} + \mathbf{z}_{(2)}\boldsymbol{\pi}_{22} + v_2 \tag{9.6}$$

where $\boldsymbol{\pi}_{21} \equiv \boldsymbol{\delta}_{(1)}/(\gamma_2 - \gamma_1)$, $\boldsymbol{\pi}_{22} = -\boldsymbol{\delta}_{(2)}/(\gamma_2 - \gamma_1)$, and $v_2 \equiv (u_1 - u_2)/(\gamma_2 - \gamma_1)$. This is the **reduced form** for y_2 because it expresses y_2 as a linear function of all of the exogenous variables and an error v_2 which, by assumption (9.2), is orthogonal to all exogenous variables: $\mathrm{E}(\mathbf{z}'v_2) = \mathbf{0}$. Importantly, the reduced form for y_2 is obtained from the two structural equations (9.4) and (9.5).

Given equation (9.4) and the reduced form (9.6), we can now use the identification condition from Chapter 5 for a linear model with a single right-hand-side endogenous variable. This condition is easy to state: the reduced form for y_2 must contain at least one exogenous variable not also in equation (9.4). This means there must be at least one element of $\mathbf{z}_{(2)}$ not in $\mathbf{z}_{(1)}$ with coefficient in equation (9.6) different from zero. Now we use the structural equations. Because $\boldsymbol{\pi}_{22}$ is proportional to $\boldsymbol{\delta}_{(2)}$, the condition is easily restated in terms of the *structural* parameters: in equation (9.5) at least one element of $\mathbf{z}_{(2)}$ not in $\mathbf{z}_{(1)}$ must have nonzero coefficient. In the supply and demand example, identification of the supply function requires at least one exogenous variable appearing in the demand function that does not also appear in the supply function; this conclusion corresponds exactly with our earlier intuition.

The condition for identifying equation (9.5) is just the mirror image: there must be at least one element of $\mathbf{z}_{(1)}$ actually appearing in equation (9.4) that is not also an element of $\mathbf{z}_{(2)}$.

Example 9.1 (Labor Supply for Married Women): Consider labor supply and demand equations for married women, with the equilibrium condition imposed:

$$hours = \gamma_1 \log(wage) + \delta_{10} + \delta_{11} educ + \delta_{12} age + \delta_{13} kids + \delta_{14} othinc + u_1$$

$$hours = \gamma_2 \log(wage) + \delta_{20} + \delta_{21} educ + \delta_{22} exper + u_2$$

The supply equation is identified because, by assumption, *exper* appears in the demand function (assuming $\delta_{22} \neq 0$) but not in the supply equation. The assumption that past experience has no direct affect on labor supply can be questioned, but it has been used by labor economists. The demand equation is identified provided that at least one of the three variables *age*, *kids*, and *othinc* actually appears in the supply equation.

We now extend this analysis to the general system (9.1). For concreteness, we study identification of the first equation:

$$y_1 = \mathbf{y}_{(1)}\boldsymbol{\gamma}_{(1)} + \mathbf{z}_{(1)}\boldsymbol{\delta}_{(1)} + u_1 = \mathbf{x}_{(1)}\boldsymbol{\beta}_{(1)} + u_1 \tag{9.7}$$

where the notation used for the subscripts is needed to distinguish an equation with exclusion restrictions from a general equation that we will study in Section 9.2.2. Assuming that the reduced forms exist, write the reduced form for $\mathbf{y}_{(1)}$ as

$$\mathbf{y}_{(1)} = \mathbf{z}\boldsymbol{\Pi}_{(1)} + \mathbf{v}_{(1)} \tag{9.8}$$

where $\mathrm{E}[\mathbf{z}'\mathbf{v}_{(1)}] = \mathbf{0}$. Further, define the $M \times M_1$ matrix *selection matrix* $\mathbf{S}_{(1)}$, which consists of zeros and ones, such that $\mathbf{z}_{(1)} = \mathbf{z}\mathbf{S}_{(1)}$. The rank condition from Chapter 5, Assumption 2SLS.2b, can be stated as

$$\operatorname{rank} \mathrm{E}[\mathbf{z}'\mathbf{x}_{(1)}] = K_1 \tag{9.9}$$

where $K_1 \equiv G_1 + M_1$. But $\mathrm{E}[\mathbf{z}'\mathbf{x}_{(1)}] = \mathrm{E}[\mathbf{z}'(\mathbf{z}\boldsymbol{\Pi}_{(1)}, \mathbf{z}\mathbf{S}_{(1)})] = \mathrm{E}(\mathbf{z}'\mathbf{z})[\boldsymbol{\Pi}_{(1)} \mid \mathbf{S}_{(1)}]$. Since we always assume that $\mathrm{E}(\mathbf{z}'\mathbf{z})$ has full rank M, assumption (9.9) is the same as

$$\operatorname{rank}[\boldsymbol{\Pi}_{(1)} \mid \mathbf{S}_{(1)}] = G_1 + M_1 \tag{9.10}$$

In other words, $[\boldsymbol{\Pi}_{(1)} \mid \mathbf{S}_{(1)}]$ must have full column rank. If the reduced form for $\mathbf{y}_{(1)}$ has been found, this condition can be checked directly. But there is one thing we can conclude immediately: because $[\boldsymbol{\Pi}_{(1)} \mid \mathbf{S}_{(1)}]$ is an $M \times (G_1 + M_1)$ matrix, a *necessary* condition for assumption (9.10) is $M \geq G_1 + M_1$, or

$$M - M_1 \geq G_1 \tag{9.11}$$

We have already encountered condition (9.11) in Chapter 5: the number of exogenous variables not appearing in the first equation, $M - M_1$, must be at least as great as the number of endogenous variables appearing on the right-hand side of the first equation, G_1. This is the **order condition** for identification of equation one. We have proven the following theorem:

THEOREM 9.1 (Order Condition with Exclusion Restrictions): In a linear system of equations with exclusion restrictions, a *necessary* condition for identifying any particular equation is that the number of excluded exogenous variables from the equation must be at least as large as the number of included right-hand-side endogenous variables in the equation.

It is important to remember that the order condition is only necessary, not sufficient, for identification. If the order condition fails for a particular equation, there is no hope of estimating the parameters in that equation. If the order condition is met, the equation *might* be identified.

9.2.2 General Linear Restrictions and Structural Equations

The identification analysis of the preceding subsection is useful when reduced forms are appended to structural equations. When an entire structural system has been specified, it is best to study identification entirely in terms of the structural parameters.

To this end, we now write the G equations in the population as

$$\begin{aligned} &\mathbf{y}\gamma_1 + \mathbf{z}\boldsymbol{\delta}_1 + u_1 = 0 \\ &\vdots \\ &\mathbf{y}\gamma_G + \mathbf{z}\boldsymbol{\delta}_G + u_G = 0 \end{aligned} \tag{9.12}$$

where $\mathbf{y} \equiv (y_1, y_2, \ldots, y_G)$ is the $1 \times G$ vector of *all* endogenous variables and $\mathbf{z} \equiv (z_1, \ldots, z_M)$ is still the $1 \times M$ vector of all exogenous variables, and probably contains unity. We maintain assumption (9.2) throughout this section and also assume that $\mathrm{E}(\mathbf{z}'\mathbf{z})$ is nonsingular. The notation here differs from that in Section 9.2.1. Here, γ_g is $G \times 1$ and $\boldsymbol{\delta}_g$ is $M \times 1$ for all $g = 1, 2, \ldots, G$, so that the system (9.12) is the general linear system without *any* restrictions on the structural parameters.

We can write this system compactly as

$$\mathbf{y}\boldsymbol{\Gamma} + \mathbf{z}\boldsymbol{\Delta} + \mathbf{u} = \mathbf{0} \tag{9.13}$$

where $\mathbf{u} \equiv (u_1, \ldots, u_G)$ is the $1 \times G$ vector of structural errors, $\mathbf{\Gamma}$ is the $G \times G$ matrix with gth column $\boldsymbol{\gamma}_g$, and $\mathbf{\Delta}$ is the $M \times G$ matrix with gth column $\boldsymbol{\delta}_g$. So that a reduced form exists, we assume that $\mathbf{\Gamma}$ is nonsingular. Let $\mathbf{\Sigma} \equiv \mathrm{E}(\mathbf{u}'\mathbf{u})$ denote the $G \times G$ variance matrix of $\mathbf{u}$, which we assume to be nonsingular. At this point, we have placed no other restrictions on $\mathbf{\Gamma}$, $\mathbf{\Delta}$, or $\mathbf{\Sigma}$.

The reduced form is easily expressed as

$$\mathbf{y} = \mathbf{z}(-\mathbf{\Delta}\mathbf{\Gamma}^{-1}) + \mathbf{u}(-\mathbf{\Gamma}^{-1}) \equiv \mathbf{z}\mathbf{\Pi} + \mathbf{v} \tag{9.14}$$

where $\mathbf{\Pi} \equiv (-\mathbf{\Delta}\mathbf{\Gamma}^{-1})$ and $\mathbf{v} \equiv \mathbf{u}(-\mathbf{\Gamma}^{-1})$. Define $\mathbf{\Lambda} \equiv \mathrm{E}(\mathbf{v}'\mathbf{v}) = \mathbf{\Gamma}^{-1\prime}\mathbf{\Sigma}\mathbf{\Gamma}^{-1}$ as the reduced form variance matrix. Because $\mathrm{E}(\mathbf{z}'\mathbf{v}) = \mathbf{0}$ and $\mathrm{E}(\mathbf{z}'\mathbf{z})$ is nonsingular, $\mathbf{\Pi}$ and $\mathbf{\Lambda}$ are identified because they can be consistently estimated given a random sample on $\mathbf{y}$ and $\mathbf{z}$ by OLS equation by equation. The question is, Under what assumptions can we recover the structural parameters $\mathbf{\Gamma}$, $\mathbf{\Delta}$, and $\mathbf{\Sigma}$ from the reduced form parameters?

It is easy to see that, without some restrictions, we will not be able to identify any of the parameters in the structural system. Let $\mathbf{F}$ be any $G \times G$ nonsingular matrix, and postmultiply equation (9.13) by $\mathbf{F}$:

$$\mathbf{y}\mathbf{\Gamma}\mathbf{F} + \mathbf{z}\mathbf{\Delta}\mathbf{F} + \mathbf{u}\mathbf{F} = \mathbf{0} \qquad \text{or} \qquad \mathbf{y}\mathbf{\Gamma}^* + \mathbf{z}\mathbf{\Delta}^* + \mathbf{u}^* = \mathbf{0} \tag{9.15}$$

where $\mathbf{\Gamma}^* \equiv \mathbf{\Gamma}\mathbf{F}$, $\mathbf{\Delta}^* \equiv \mathbf{\Delta}\mathbf{F}$, and $\mathbf{u}^* \equiv \mathbf{u}\mathbf{F}$; note that $\mathrm{Var}(\mathbf{u}^*) = \mathbf{F}'\mathbf{\Sigma}\mathbf{F}$. Simple algebra shows that equations (9.15) and (9.13) have *identical* reduced forms. This result means that, without restrictions on the structural parameters, there are many **equivalent structures** in the sense that they lead to the same reduced form. In fact, there is an equivalent structure for each nonsingular $\mathbf{F}$.

Let $\mathbf{B} \equiv \begin{pmatrix} \mathbf{\Gamma} \\ \mathbf{\Delta} \end{pmatrix}$ be the $(G + M) \times G$ matrix of structural parameters in equation (9.13). If $\mathbf{F}$ is any nonsingular $G \times G$ matrix, then $\mathbf{F}$ represents an **admissible linear transformation** if

1. $\mathbf{BF}$ satisfies all restrictions on $\mathbf{B}$.
2. $\mathbf{F}'\mathbf{\Sigma}\mathbf{F}$ satisfies all restrictions on $\mathbf{\Sigma}$.

To identify the system, we need enough prior information on the structural parameters $(\mathbf{B}, \mathbf{\Sigma})$ so that $\mathbf{F} = \mathbf{I}_G$ is the only admissible linear transformation.

In most applications identification of $\mathbf{B}$ is of primary interest, and this identification is achieved by putting restrictions directly on $\mathbf{B}$. As we will touch on in Section 9.4.2, it is possible to put restrictions on $\mathbf{\Sigma}$ in order to identify $\mathbf{B}$, but this approach is somewhat rare in practice. Until we come to Section 9.4.2, $\mathbf{\Sigma}$ is an unrestricted $G \times G$ positive definite matrix.

As before, we consider identification of the first equation:

$$\mathbf{y}\boldsymbol{\gamma}_1 + \mathbf{z}\boldsymbol{\delta}_1 + u_1 = 0 \tag{9.16}$$

or $\gamma_{11}y_1 + \gamma_{12}y_2 + \cdots + \gamma_{1G}y_G + \delta_{11}z_1 + \delta_{12}z_2 + \cdots + \delta_{1M}z_M + u_1 = 0$. The first restriction we make on the parameters in equation (9.16) is the **normalization restriction** that one element of $\boldsymbol{\gamma}_1$ is -1. Each equation in the system (9.1) has a normalization restriction because one variable is taken to be the left-hand-side explained variable. In applications, there is usually a natural normalization for each equation. If there is not, we should ask whether the system satisfies the autonomy requirement discussed in Section 9.1. (Even in models that satisfy the autonomy requirement, we often have to choose between reasonable normalization conditions. For example, in Example 9.1, we could have specified the second equation to be a wage offer equation rather than a labor demand equation.)

Let $\boldsymbol{\beta}_1 \equiv (\boldsymbol{\gamma}_1', \boldsymbol{\delta}_1')'$ be the $(G+M) \times 1$ vector of structural parameters in the first equation. With a normalization restriction there are $(G+M) - 1$ unknown elements in $\boldsymbol{\beta}_1$. Assume that prior knowledge about $\boldsymbol{\beta}_1$ can be expressed as

$$\mathbf{R}_1\boldsymbol{\beta}_1 = \mathbf{0} \tag{9.17}$$

where $\mathbf{R}_1$ is a $J_1 \times (G+M)$ matrix of known constants, and J_1 is the number of restrictions on $\boldsymbol{\beta}_1$ (in addition to the normalization restriction). We assume that rank $\mathbf{R}_1 = J_1$, so that there are no redundant restrictions. The restrictions in assumption (9.17) are sometimes called **homogeneous linear restrictions**, but, when coupled with a normalization assumption, equation (9.17) actually allows for nonhomogeneous restrictions.

Example 9.2 (A Three-Equation System): Consider the first equation in a system with $G = 3$ and $M = 4$:

$$y_1 = \gamma_{12}y_2 + \gamma_{13}y_3 + \delta_{11}z_1 + \delta_{12}z_2 + \delta_{13}z_3 + \delta_{14}z_4 + u_1$$

so that $\boldsymbol{\gamma}_1 = (-1, \gamma_{12}, \gamma_{13})'$, $\boldsymbol{\delta}_1 = (\delta_{11}, \delta_{12}, \delta_{13}, \delta_{14})'$, and $\boldsymbol{\beta}_1 = (-1, \gamma_{12}, \gamma_{13}, \delta_{11}, \delta_{12}, \delta_{13}, \delta_{14})'$. (We can set $z_1 = 1$ to allow an intercept.) Suppose the restrictions on the structural parameters are $\gamma_{12} = 0$ and $\delta_{13} + \delta_{14} = 3$. Then $J_1 = 2$ and

$$\mathbf{R}_1 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Straightforward multiplication gives $\mathbf{R}_1\boldsymbol{\beta}_1 = (\gamma_{12}, \delta_{13} + \delta_{14} - 3)'$, and setting this vector to zero as in equation (9.17) incorporates the restrictions on $\boldsymbol{\beta}_1$.

Given the linear restrictions in equation (9.17), when are these and the normalization restriction enough to identify $\boldsymbol{\beta}_1$? Let $\mathbf{F}$ again be any $G \times G$ nonsingular matrix, and write it in terms of its columns as $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_G)$. Define a linear transformation of $\mathbf{B}$ as $\mathbf{B}^* = \mathbf{BF}$, so that the first column of $\mathbf{B}^*$ is $\boldsymbol{\beta}_1^* \equiv \mathbf{Bf}_1$. We need to find a condition so that equation (9.17) allows us to distinguish $\boldsymbol{\beta}_1$ from any other $\boldsymbol{\beta}_1^*$. For the moment, ignore the normalization condition. The vector $\boldsymbol{\beta}_1^*$ satisfies the linear restrictions embodied by $\mathbf{R}_1$ if and only if

$$\mathbf{R}_1\boldsymbol{\beta}_1^* = \mathbf{R}_1(\mathbf{Bf}_1) = (\mathbf{R}_1\mathbf{B})\mathbf{f}_1 = \mathbf{0} \tag{9.18}$$

Naturally, $(\mathbf{R}_1\mathbf{B})\mathbf{f}_1 = \mathbf{0}$ is true for $\mathbf{f}_1 = \mathbf{e}_1 \equiv (1, 0, 0, \ldots, 0)'$, since then $\boldsymbol{\beta}_1^* = \mathbf{Bf}_1 = \boldsymbol{\beta}_1$. Since assumption (9.18) holds for $\mathbf{f}_1 = \mathbf{e}_1$ it clearly holds for any scalar multiple of $\mathbf{e}_1$. The key to identification is that vectors of the form $c_1\mathbf{e}_1$, for some constant c_1, are the *only* vectors $\mathbf{f}_1$ satisfying condition (9.18). If condition (9.18) holds for vectors $\mathbf{f}_1$ other than scalar multiples of $\mathbf{e}_1$ then we have no hope of identifying $\boldsymbol{\beta}_1$.

Stating that condition (9.18) holds only for vectors of the form $c_1\mathbf{e}_1$ just means that the null space of $\mathbf{R}_1\mathbf{B}$ has dimension unity. Equivalently, because $\mathbf{R}_1\mathbf{B}$ has G columns,

$$\text{rank } \mathbf{R}_1\mathbf{B} = G - 1 \tag{9.19}$$

This is the **rank condition** for identification of $\boldsymbol{\beta}_1$ in the first structural equation under general linear restrictions. Once condition (9.19) is known to hold, the normalization restriction allows us to distinguish $\boldsymbol{\beta}_1$ from any other scalar multiple of $\boldsymbol{\beta}_1$.

THEOREM 9.2 (Rank Condition for Identification): Let $\boldsymbol{\beta}_1$ be the $(G + M) \times 1$ vector of structural parameters in the first equation, with the normalization restriction that one of the coefficients on an endogenous variable is -1. Let the additional information on $\boldsymbol{\beta}_1$ be given by restriction (9.17). Then $\boldsymbol{\beta}_1$ is identified if and only if the rank condition (9.19) holds.

As promised earlier, the rank condition in this subsection depends on the *structural* parameters, $\mathbf{B}$. We can determine whether the first equation is identified by studying the matrix $\mathbf{R}_1\mathbf{B}$. Since this matrix can depend on *all* structural parameters, we must generally specify the entire structural model.

The $J_1 \times G$ matrix $\mathbf{R}_1\mathbf{B}$ can be written as $\mathbf{R}_1\mathbf{B} = [\mathbf{R}_1\boldsymbol{\beta}_1, \mathbf{R}_1\boldsymbol{\beta}_2, \ldots, \mathbf{R}_1\boldsymbol{\beta}_G]$, where $\boldsymbol{\beta}_g$ is the $(G + M) \times 1$ vector of structural parameters in equation g. By assumption (9.17), the first column of $\mathbf{R}_1\mathbf{B}$ is the zero vector. Therefore, $\mathbf{R}_1\mathbf{B}$ cannot have rank larger than $G - 1$. What we must check is whether the columns of $\mathbf{R}_1\mathbf{B}$ other than the first form a linearly independent set.

Using condition (9.19) we can get a more general form of the order condition. Because $\boldsymbol{\Gamma}$ is nonsingular, $\mathbf{B}$ necessarily has rank G (full column rank). Therefore, for condition (9.19) to hold, we must have rank $\mathbf{R}_1 \geq G - 1$. But we have assumed that rank $\mathbf{R}_1 = J_1$, which is the row dimension of $\mathbf{R}_1$.

THEOREM 9.3 (Order Condition for Identification): In system (9.12) under assumption (9.17), a *necessary* condition for the first equation to be identified is

$$J_1 \geq G - 1 \tag{9.20}$$

where J_1 is the row dimension of $\mathbf{R}_1$. Equation (9.20) is the general form of the order condition.

We can summarize the steps for checking whether the first equation in the system is identified.

1. Set one element of $\boldsymbol{\gamma}_1$ to -1 as a normalization.
2. Define the $J_1 \times (G + M)$ matrix $\mathbf{R}_1$ such that equation (9.17) captures all restrictions on $\boldsymbol{\beta}_1$.
3. If $J_1 < G - 1$, the first equation is not identified.
4. If $J_1 \geq G - 1$, the equation might be identified. Let $\mathbf{B}$ be the matrix of all structural parameters with only the normalization restrictions imposed, and compute $\mathbf{R}_1\mathbf{B}$. Now impose the restrictions in the entire system and check the rank condition (9.19).

The simplicity of the order condition makes it attractive as a tool for studying identification. Nevertheless, it is not difficult to write down examples where the order condition is satisfied but the rank condition fails.

Example 9.3 (Failure of the Rank Condition): Consider the following three-equation structural model in the population ($G = 3, M = 4$):

$$y_1 = \gamma_{12} y_2 + \gamma_{13} y_3 + \delta_{11} z_1 + \delta_{13} z_3 + u_1 \tag{9.21}$$

$$y_2 = \gamma_{21} y_1 + \delta_{21} z_1 + u_2 \tag{9.22}$$

$$y_3 = \delta_{31} z_1 + \delta_{32} z_2 + \delta_{33} z_3 + \delta_{34} z_4 + u_3 \tag{9.23}$$

where $z_1 \equiv 1$, $\mathrm{E}(u_g) = 0$, $g = 1, 2, 3$, and each z_j is uncorrelated with each u_g. Note that the third equation is already a reduced form equation (although it may also have a structural interpretation). In equation (9.21) we have set $\gamma_{11} = -1$, $\delta_{12} = 0$, and $\delta_{14} = 0$. Since this equation contains two right-hand-side endogenous variables and there are two excluded exogenous variables, it passes the order condition.

To check the rank condition, let $\boldsymbol{\beta}_1$ denote the 7×1 vector of parameters in the first equation with only the normalization restriction imposed: $\boldsymbol{\beta}_1 = (-1, \gamma_{12}, \gamma_{13}, \delta_{11}, \delta_{12}, \delta_{13}, \delta_{14})'$. The restrictions $\delta_{12} = 0$ and $\delta_{14} = 0$ are obtained by choosing

$$\mathbf{R}_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Let $\mathbf{B}$ be the full 7×3 matrix of parameters with only the three normalizations imposed [so that $\boldsymbol{\beta}_2 = (\gamma_{21}, -1, \gamma_{23}, \delta_{21}, \delta_{22}, \delta_{23}, \delta_{24})'$ and $\boldsymbol{\beta}_3 = (\gamma_{31}, \gamma_{32}, -1, \delta_{31}, \delta_{32}, \delta_{33}, \delta_{34})'$]. Matrix multiplication gives

$$\mathbf{R}_1\mathbf{B} = \begin{pmatrix} \delta_{12} & \delta_{22} & \delta_{32} \\ \delta_{14} & \delta_{24} & \delta_{34} \end{pmatrix}$$

Now we impose all of the restrictions in the system. In addition to the restrictions $\delta_{12} = 0$ and $\delta_{14} = 0$ from equation (9.21), we also have $\delta_{22} = 0$ and $\delta_{24} = 0$ from equation (9.22). Therefore, with all restrictions imposed,

$$\mathbf{R}_1\mathbf{B} = \begin{pmatrix} 0 & 0 & \delta_{32} \\ 0 & 0 & \delta_{34} \end{pmatrix} \tag{9.24}$$

The rank of this matrix is at most unity, and so the rank condition fails because $G - 1 = 2$.

Equation (9.22) easily passes the order condition. It is left to you to show that the rank condition holds if and only if $\delta_{13} \neq 0$ and at least one of δ_{32} and δ_{34} is different from zero. The third equation is identified because it contains no endogenous explanatory variables.

When the restrictions on $\boldsymbol{\beta}_1$ consist entirely of normalization and exclusion restrictions, the order condition (9.20) reduces to the order condition (9.11), as can be seen by the following argument. When all restrictions are exclusion restrictions, the matrix $\mathbf{R}_1$ consists only of zeros and ones, and the number of rows in $\mathbf{R}_1$ equals the number of excluded right-hand-side endogenous variables, $G - G_1 - 1$, plus the number of excluded exogenous variables, $M - M_1$. In other words, $J_1 = (G - G_1 - 1) + (M - M_1)$, and so the order condition (9.20) becomes $(G - G_1 - 1) + (M - M_1) \geq G - 1$, which, upon rearrangement, becomes condition (9.11).

9.2.3 Unidentified, Just Identified, and Overidentified Equations

We have seen that, for identifying a single equation the rank condition (9.19) is necessary and sufficient. When condition (9.19) fails, we say that the equation is **unidentified**.

When the rank condition holds, it is useful to refine the sense in which the equation is identified. If $J_1 = G - 1$, then we have just enough identifying information. If we were to drop one restriction in $\mathbf{R}_1$, we would necessarily lose identification of the first equation because the order condition would fail. Therefore, when $J_1 = G - 1$, we say that the equation is **just identified**.

If $J_1 > G - 1$, it is often possible to drop one or more restrictions on the parameters of the first equation and still achieve identification. In this case we say the equation is **overidentified**. Necessary but not sufficient for overidentification is $J_1 > G - 1$. It is possible that J_1 is strictly greater than $G - 1$ but the restrictions are such that dropping one restriction loses identification, in which case the equation is not overidentified.

In practice, we often appeal to the order condition to determine the degree of overidentification. While in special circumstances this approach can fail to be accurate, for most applications it is reasonable. Thus, for the first equation, $J_1 - (G - 1)$ is usually intepreted as the number of **overidentifying restrictions**.

Example 9.4 (Overidentifying Restrictions): Consider the two-equation system

$$y_1 = \gamma_{12} y_2 + \delta_{11} z_1 + \delta_{12} z_2 + \delta_{13} z_3 + \delta_{14} z_4 + u_1 \tag{9.25}$$

$$y_2 = \gamma_{21} y_1 + \delta_{21} z_1 + \delta_{22} z_2 + u_2 \tag{9.26}$$

where $\mathrm{E}(z_j u_g) = 0$, all j and g. Without further restrictions, equation (9.25) fails the order condition because every exogenous variable appears on the right-hand side, and the equation contains an endogenous variable. Using the order condition, equation (9.26) is overidentified, with one overidentifying restriction. If z_3 does not actually appear in equation (9.25), then equation (9.26) is just identified, assuming that $\delta_{14} \neq 0$.

9.3 Estimation after Identification

9.3.1 The Robustness-Efficiency Trade-off

All SEMs with linearly homogeneous restrictions within each equation can be written with exclusion restrictions as in the system (9.1); doing so may require redefining some of the variables. If we let $\mathbf{x}_{(g)} = (\mathbf{y}_{(g)}, \mathbf{z}_{(g)})$ and $\boldsymbol{\beta}_{(g)} = (\boldsymbol{\gamma}'_{(g)}, \boldsymbol{\delta}'_{(g)})'$, then the system (9.1) is in the general form (8.11) with the slight change in notation. Under assumption (9.2) the matrix of instruments for observation i is the $G \times GM$ matrix

$$\mathbf{Z}_i \equiv \mathbf{I}_G \otimes \mathbf{z}_i \tag{9.27}$$

If every equation in the system passes the rank condition, a system estimation procedure—such as 3SLS or the more general minimum chi-square estimator—can be used. Alternatively, the equations of interest can be estimated by 2SLS. The bottom line is that the methods studied in Chapters 5 and 8 are directly applicable. All of the tests we have covered apply, including the tests of overidentifying restrictions in Chapters 6 and 8, and the single-equation tests for endogeneity in Chapter 6.

When estimating a simultaneous equations system, it is important to remember the pros and cons of full system estimation. If all equations are correctly specified, system procedures are asymptotically more efficient than a single-equation procedure such as 2SLS. But single-equation methods are more robust. If interest lies, say, in the first equation of a system, 2SLS is consistent and asymptotically normal provided the first equation is correctly specified and the instruments are exogenous. However, if one equation in a system is misspecified, the 3SLS or GMM estimates of all the parameters are generally inconsistent.

Example 9.5 (Labor Supply for Married, Working Women): Using the data in MROZ.RAW, we estimate a labor supply function for working, married women. Rather than specify a demand function, we specify the second equation as a wage offer function and impose the equilibrium condition:

$$\begin{aligned} hours = {} & \gamma_{12}\log(wage) + \delta_{10} + \delta_{11} educ + \delta_{12} age + \delta_{13} kidslt6 \\ & + \delta_{14} kidsge6 + \delta_{15} nwifeinc + u_1 \end{aligned} \tag{9.28}$$

$$\log(wage) = \gamma_{21} hours + \delta_{20} + \delta_{21} educ + \delta_{22} exper + \delta_{23} exper^2 + u_2 \tag{9.29}$$

where *kidslt6* is number of children less than 6, *kidsge6* is number of children between 6 and 18, and *nwifeinc* is income other than the woman's labor income. We assume that u_1 and u_2 have zero mean conditional on *educ*, *age*, *kidslt6*, *kidsge6*, *nwifeinc*, and *exper*.

The key restriction on the labor supply function is that *exper* (and $exper^2$) have no direct effect on current annual hours. This identifies the labor supply function with one overidentifying restriction, as used by Mroz (1987). We estimate the labor supply function first by OLS [to see what ignoring the endogeneity of log(*wage*) does] and then by 2SLS, using as instruments all exogenous variables in equations (9.28) and (9.29).

There are 428 women who worked at some time during the survey year, 1975. The average annual hours are about 1,303 with a minimum of 12 and a maximum of 4,950.

We first estimate the labor supply function by OLS:

$$\begin{aligned} \widehat{hours} = {} & \underset{(340.1)}{2{,}114.7} - \underset{(54.22)}{17.41}\,\log(wage) - \underset{(17.97)}{14.44}\,educ - \underset{(5.53)}{7.73}\,age \\ & - \underset{(100.01)}{342.50}\,kidslt6 - \underset{(30.83)}{115.02}\,kidsge6 - \underset{(3.66)}{4.35}\,nwifeinc \end{aligned}$$

The OLS estimates indicate a downward-sloping labor supply function, although the estimate on log(*wage*) is statistically insignificant.

The estimates are much different when we use 2SLS:

$$\begin{aligned}\widehat{hours} = {} & \underset{(594.2)}{2{,}432.2} + \underset{(480.74)}{1{,}544.82}\ \log(wage) - \underset{(58.14)}{177.45}\ educ - \underset{(9.58)}{10.78}\ age \\ & - \underset{(176.93)}{210.83}\ kidslt6 - \underset{(56.92)}{47.56}\ kidsge6 - \underset{(6.48)}{9.25}\ nwifeinc\end{aligned}$$

The estimated labor supply elasticity is $1{,}544.82/hours$. At the mean *hours* for working women, 1,303, the estimated elasticity is about 1.2, which is quite large.

The supply equation has a single overidentifying restriction. The regression of the 2SLS residuals $\hat{u}_1$ on all exogenous variables produces $R_u^2 = .002$, and so the test statistic is $428(.002) \approx .856$ with p-value $\approx .355$; the overidentifying restriction is not rejected.

Under the exclusion restrictions we have imposed, the wage offer function (9.29) is also identified. Before estimating the equation by 2SLS, we first estimate the reduced form for *hours* to ensure that the exogenous variables excluded from equation (9.29) are jointly significant. The p-value for the F test of joint significance of *age*, *kidslt6*, *kidsge6*, and *nwifeinc* is about .0009. Therefore, we can proceed with 2SLS estimation of the wage offer equation. The coefficient on *hours* is about .00016 (standard error $\approx$ .00022), and so the wage offer does not appear to differ by hours worked. The remaining coefficients are similar to what is obtained by dropping *hours* from equation (9.29) and estimating the equation by OLS. (For example, the 2SLS coefficient on education is about .111 with se $\approx$ .015.)

Interestingly, while the wage offer function (9.29) is identified, the analogous labor demand function is apparently unidentified. (This finding shows that choosing the normalization—that is, choosing between a labor demand function and a wage offer function—is not innocuous.) The labor demand function, written in equilibrium, would look like this:

$$hours = \gamma_{22} \log(wage) + \delta_{20} + \delta_{21} educ + \delta_{22} exper + \delta_{23} exper^2 + u_2 \tag{9.30}$$

Estimating the reduced form for log(*wage*) and testing for joint significance of *age*, *kidslt6*, *kidsge6*, and *nwifeinc* yields a p-value of about .46, and so the exogenous variables excluded from equation (9.30) would not seem to appear in the reduced form for log(*wage*). Estimation of equation (9.30) by 2SLS would be pointless. [You are invited to estimate equation (9.30) by 2SLS to see what happens.]

It would be more efficient to estimate equations (9.28) and (9.29) by 3SLS, since each equation is overidentified (assuming the homoskedasticity assumption SIV.5). If heteroskedasticity is suspected, we could use the general minimum chi-square estimator. A system procedure is more efficient for estimating the labor supply function because it uses the information that *age*, *kidslt6*, *kidsge6*, and *nwifeinc* do not appear in the log(*wage*) equation. If these exclusion restrictions are wrong, the 3SLS estimators of parameters in *both* equations are generally inconsistent. Problem 9.9 asks you to obtain the 3SLS estimates for this example.

9.3.2 When Are 2SLS and 3SLS Equivalent?

In Section 8.4 we discussed the relationship between 2SLS and 3SLS for a general linear system. Applying that discussion to linear SEMs, we can immediately draw the following conclusions: (1) if each equation is just identified, 2SLS equation by equation is algebraically identical to 3SLS, which is the same as the IV estimator in equation (8.22); (2) regardless of the degree of overidentification, 2SLS equation by equation and 3SLS are identical if $\hat{\boldsymbol{\Sigma}}$ is diagonal.

Another useful equivalence result in the context of linear SEMs is as follows. Suppose that the first equation in a system is overidentified but every other equation is just identified. (A special case occurs when the first equation is a structural equation and all remaining equations are unrestricted reduced forms.) Then the 2SLS estimator of the first equation is the same as the 3SLS estimator. This result follows as a special case of Schmidt (1976, Theorem 5.2.13).

9.3.3 Estimating the Reduced Form Parameters

So far, we have discussed estimation of the structural parameters. The usual justifications for focusing on the structural parameters are as follows: (1) we are interested in estimates of "economic parameters" (such as labor supply elasticities) for curiosity's sake; (2) estimates of structural parameters allow us to obtain the effects of a variety of policy interventions (such as changes in tax rates); and (3) even if we want to estimate the reduced form parameters, we often can do so more efficiently by first estimating the structural parameters. Concerning the second reason, if the goal is to estimate, say, the equilibrium change in hours worked given an exogenous change in a marginal tax rate, we must ultimately estimate the reduced form.

As another example, we might want to estimate the effect on county-level alcohol consumption due to an increase in exogenous alcohol taxes. In other words, we are interested in $\partial \mathrm{E}(y_g \mid \mathbf{z})/\partial z_j = \pi_{gj}$, where y_g is alcohol consumption and z_j is the tax on alcohol. Under weak assumptions, reduced form equations exist, and each equation of the reduced form can be estimated by ordinary least squares. Without placing any restrictions on the reduced form, OLS equation by equation is identical to SUR estimation (see Section 7.7). In other words, we do not need to analyze the structural equations at all in order to consistently estimate the reduced form parameters. Ordinary least squares estimates of the reduced form parameters are robust in the sense that they do not rely on any identification assumptions imposed on the structural system.

If the structural model is correctly specified and at least one equation is overidentified, we obtain asymptotically more efficient estimators of the reduced form parameters by deriving the estimates from the structural parameter estimates. In particular, given the structural parameter estimates $\hat{\mathbf{\Delta}}$ and $\hat{\mathbf{\Gamma}}$, we can obtain the reduced form estimates as $\hat{\mathbf{\Pi}} = -\hat{\mathbf{\Delta}}\hat{\mathbf{\Gamma}}^{-1}$ [see equation (9.14)]. These are consistent, $\sqrt{N}$-asymptotically normal estimators (although the asymptotic variance matrix is somewhat complicated). From Problem 3.9, we obtain the most efficient estimator of $\mathbf{\Pi}$ by using the most efficient estimators of $\mathbf{\Delta}$ and $\mathbf{\Gamma}$ (minimum chi-square or, under system homoskedasticity, 3SLS).

Just as in estimating the structural parameters, there is a robustness-efficiency trade-off in estimating the π_{gj}. As mentioned earlier, the OLS estimators of each reduced form are robust to misspecification of any restrictions on the structural equations (although, as always, each element of $\mathbf{z}$ should be exogenous for OLS to be consistent). The estimators of the π_{gj} derived from estimators of $\mathbf{\Delta}$ and $\mathbf{\Gamma}$—whether the latter are 2SLS or system estimators—are generally nonrobust to incorrect restrictions on the structural system. See Problem 9.11 for a simple illustration.

9.4 Additional Topics in Linear SEMs

9.4.1 Using Cross Equation Restrictions to Achieve Identification

So far we have discussed identification of a single equation using only within-equation parameter restrictions [see assumption (9.17)]. This is by far the leading case, especially when the system represents a simultaneous equations model with truly autonomous equations. Nevertheless, occasionally economic theory implies parameter restrictions across different equations in a system that contains endogenous variables. Not surprisingly, such **cross equation restrictions** are generally useful for identifying equations. A general treatment is beyond the scope of our analysis. Here we just give an example to show how identification and estimation work.

Consider the two-equation system

$$y_1 = \gamma_{12} y_2 + \delta_{11} z_1 + \delta_{12} z_2 + \delta_{13} z_3 + u_1 \tag{9.31}$$

$$y_2 = \gamma_{21} y_1 + \delta_{21} z_1 + \delta_{22} z_2 + u_2 \tag{9.32}$$

where each z_j is uncorrelated with u_1 and u_2 (z_1 can be unity to allow for an intercept). Without further information, equation (9.31) is unidentified, and equation (9.32) is just identified if and only if $\delta_{13} \neq 0$. We maintain these assumptions in what follows.

Now suppose that $\delta_{12} = \delta_{22}$. Because δ_{22} is identified in equation (9.32) we can treat it as known for studying identification of equation (9.31). But $\delta_{12} = \delta_{22}$, and so we can write

$$y_1 - \delta_{12} z_2 = \gamma_{12} y_2 + \delta_{11} z_1 + \delta_{13} z_3 + u_1 \tag{9.33}$$

where $y_1 - \delta_{12} z_2$ is effectively known. Now the right-hand side of equation (9.33) has one endogenous variable, y_2, and the two exogenous variables z_1 and z_3. Because z_2 is excluded from the right-hand side, we can use z_2 as an instrument for y_2, as long as z_2 appears in the reduced form for y_2. This is the case provided $\delta_{12} = \delta_{22} \neq 0$.

This approach to showing that equation (9.31) is identified also suggests a consistent estimation procedure: first, estimate equation (9.32) by 2SLS using (z_1, z_2, z_3) as instruments, and let $\hat{\delta}_{22}$ be the estimator of δ_{22}. Then, estimate

$$y_1 - \hat{\delta}_{22} z_2 = \gamma_{12} y_2 + \delta_{11} z_1 + \delta_{13} z_3 + error$$

by 2SLS using (z_1, z_2, z_3) as instruments. Since $\hat{\delta}_{22} \xrightarrow{p} \delta_{12}$ when $\delta_{12} = \delta_{22} \neq 0$, this last step produces consistent estimators of γ_{12}, δ_{11}, and δ_{13}. Unfortunately, the usual 2SLS standard errors obtained from the final estimation would not be valid because of the preliminary estimation of δ_{22}.

It is easier to use a system procedure when cross equation restrictions are present because the asymptotic variance can be obtained directly. We can always rewrite the system in a linear form with the restrictions imposed. For this example, one way to do so is to write the system as

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_2 & z_1 & z_2 & z_3 & 0 & 0 \\ 0 & 0 & z_2 & 0 & y_1 & z_1 \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \tag{9.34}$$

where $\boldsymbol{\beta} = (\gamma_{12}, \delta_{11}, \delta_{12}, \delta_{13}, \gamma_{21}, \delta_{21})'$. The parameter δ_{22} does not show up in $\boldsymbol{\beta}$ because we have imposed the restriction $\delta_{12} = \delta_{22}$ by appropriate choice of the matrix of explanatory variables.

The matrix of instruments is $\mathbf{I}_2 \otimes \mathbf{z}$, meaning that we just use all exogenous variables as instruments in each equation. Since $\mathbf{I}_2 \otimes \mathbf{z}$ has six columns, the order condition is exactly satisfied (there are six elements of $\boldsymbol{\beta}$), and we have already seen when the rank condition holds. The system can be consistently estimated using GMM or 3SLS.

9.4.2 Using Covariance Restrictions to Achieve Identification

In most applications of linear SEMs, identification is obtained by putting restrictions on the matrix of structural parameters $\mathbf{B}$. Occasionally, we are willing to put restrictions on the variance matrix $\boldsymbol{\Sigma}$ of the structural errors. Such restrictions, which are almost always zero covariance assumptions, can help identify the structural parameters in some equations. For general treatments see Hausman (1983) and Hausman, Newey, and Taylor (1987). We give a couple of examples to show how identification with covariance restrictions works.

The first example is the two-equation system

$$y_1 = \gamma_{12} y_2 + \delta_{11} z_1 + \delta_{13} z_3 + u_1 \tag{9.35}$$

$$y_2 = \gamma_{21} y_1 + \delta_{21} z_1 + \delta_{22} z_2 + \delta_{23} z_3 + u_2 \tag{9.36}$$

Equation (9.35) is just identified if $\delta_{22} \neq 0$, which we assume, while equation (9.36) is unidentified without more information. Suppose that we have one piece of additional information in terms of a covariance restriction:

$$\mathrm{Cov}(u_1, u_2) = \mathrm{E}(u_1 u_2) = 0 \tag{9.37}$$

In other words, if $\boldsymbol{\Sigma}$ is the 2×2 structural variance matrix, we are assuming that $\boldsymbol{\Sigma}$ is diagonal. Assumption (9.37), along with $\delta_{22} \neq 0$, is enough to identify equation (9.36).

Here is a simple way to see how assumption (9.37) identifies equation (9.36). First, because γ_{12}, δ_{11}, and δ_{13} are identified, we can treat them as known when studying identification of equation (9.36). But if the parameters in equation (9.35) are known, u_1 is effectively known. By assumption (9.37), u_1 is uncorrelated with u_2, and u_1 is certainly partially correlated with y_1. Thus, we effectively have (z_1, z_2, z_3, u_1) as instruments available for estimating equation (9.36), and this result shows that equation (9.36) is identified.

We can use this method for verifying identification to obtain consistent estimators. First, estimate equation (9.35) by 2SLS using instruments (z_1, z_2, z_3) and save the 2SLS residuals, $\hat{u}_1$. Then estimate equation (9.36) by 2SLS using instruments $(z_1, z_2, z_3, \hat{u}_1)$. The fact that $\hat{u}_1$ depends on estimates from a prior stage does not affect consistency. But inference is complicated because of the estimation of u_1: condition (6.8) does not hold because u_1 depends on y_2, which is correlated with u_2.

The most efficient way to use covariance restrictions is to write the entire set of orthogonality conditions as $\mathrm{E}[\mathbf{z}' u_1(\boldsymbol{\beta}_1)] = \mathbf{0}$, $\mathrm{E}[\mathbf{z}' u_2(\boldsymbol{\beta}_2)] = \mathbf{0}$, and

$$\mathrm{E}[u_1(\boldsymbol{\beta}_1) u_2(\boldsymbol{\beta}_2)] = 0 \tag{9.38}$$

where the notation $u_1(\boldsymbol{\beta}_1)$ emphasizes that the errors are functions of the structural parameters $\boldsymbol{\beta}_1$—with normalization and exclusion restrictions imposed—and similarly for $u_2(\boldsymbol{\beta}_2)$. For example, from equation (9.35), $u_1(\boldsymbol{\beta}_1) = y_1 - \gamma_{12}y_2 - \delta_{11}z_1 - \delta_{13}z_3$. Equation (9.38), because it is nonlinear in $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, takes us outside the realm of linear moment restrictions. In Chapter 14 we will use nonlinear moment conditions in GMM estimation.

A general example with covariance restrictions is a **fully recursive system**. First, a **recursive system** can be written as

$$\begin{aligned} y_1 &= \mathbf{z}\boldsymbol{\delta}_1 + u_1 \\ y_2 &= \gamma_{21}y_1 + \mathbf{z}\boldsymbol{\delta}_2 + u_2 \\ y_3 &= \gamma_{31}y_1 + \gamma_{32}y_2 + \mathbf{z}\boldsymbol{\delta}_3 + u_3 \\ &\vdots \\ y_G &= \gamma_{G1}y_1 + \cdots + \gamma_{G,G-1}y_{G-1} + \mathbf{z}\boldsymbol{\delta}_G + u_G \end{aligned} \tag{9.39}$$

so that in each equation only endogenous variables from previous equations appear on the right-hand side. We have allowed all exogenous variables to appear in each equation, and we maintain assumption (9.2).

The first equation in the system (9.39) is clearly identified and can be estimated by OLS. Without further exclusion restrictions none of the remaining equations is identified, but each is identified if we assume that the structural errors are pairwise uncorrelated:

$$\mathrm{Cov}(u_g, u_h) = 0, \qquad g \neq h \tag{9.40}$$

This assumption means that $\boldsymbol{\Sigma}$ is a $G \times G$ diagonal matrix. Equations (9.39) and (9.40) define a fully recursive system. Under these assumptions, the right-hand-side variables in equation g are each uncorrelated with u_g; this fact is easily seen by starting with the first equation and noting that y_1 is a linear function of $\mathbf{z}$ and u_1. Then, in the second equation, y_1 is uncorrelated with u_2 under assumption (9.40). But y_2 is a linear function of $\mathbf{z}$, u_1, and u_2, and so y_2 and y_1 are both uncorrelated with u_3 in the third equation. And so on. It follows that each equation in the system is consistently estimated by ordinary least squares.

It turns out that OLS equation by equation is not necessarily the most efficient estimator in fully recursive systems, even though $\boldsymbol{\Sigma}$ is a diagonal matrix. Generally, efficiency can be improved by adding the zero covariance restrictions to the orthogonality conditions, as in equation (9.38), and applying nonlinear GMM estimation. See Lahiri and Schmidt (1978) and Hausman, Newey, and Taylor (1987).

9.4.3 Subtleties Concerning Identification and Efficiency in Linear Systems

So far we have discussed identification and estimation under the assumption that each exogenous variable appearing in the system, z_j, is *uncorrelated* with each structural error, u_g. It is important to assume only zero correlation in the general treatment because we often add a reduced form equation for an endogenous variable to a structural system, and zero correlation is all we should impose in linear reduced forms.

For entirely structural systems, it is often natural to assume that the structural errors satisfy the zero conditional mean assumption

$$\mathrm{E}(u_g \,|\, \mathbf{z}) = 0, \qquad g = 1, 2, \ldots, G \tag{9.41}$$

In addition to giving the parameters in the structural equations the appropriate partial effect interpretations, assumption (9.41) has some interesting statistical implications: *any* function of $\mathbf{z}$ is uncorrelated with each error u_g. Therefore, in the labor supply example (9.28), age^2, $\log(age)$, $educ{\cdot}exper$, and so on (there are too many functions to list) are all uncorrelated with u_1 and u_2. Realizing this fact, we might ask, Why not use nonlinear functions of $\mathbf{z}$ as additional instruments in estimation?

We need to break the answer to this question into two parts. The first concerns identification, and the second concerns efficiency. For identification, the bottom line is this: adding nonlinear functions of $\mathbf{z}$ to the instrument list *cannot* help with identification in linear systems. You were asked to show this generally in Problem 8.4, but the main points can be illustrated with a simple model:

$$y_1 = \gamma_{12} y_2 + \delta_{11} z_1 + \delta_{12} z_2 + u_1 \tag{9.42}$$

$$y_2 = \gamma_{21} y_1 + \delta_{21} z_1 + u_2 \tag{9.43}$$

$$\mathrm{E}(u_1 | \mathbf{z}) = \mathrm{E}(u_2 \,|\, \mathbf{z}) = 0 \tag{9.44}$$

From the order condition in Section 9.2.2, equation (9.42) is not identified, and equation (9.43) is identified if and only if $\delta_{12} \neq 0$. Knowing properties of conditional expectations, we might try something clever to identify equation (9.42): since, say, z_1^2 is uncorrelated with u_1 under assumption (9.41), and z_1^2 would appear to be correlated with y_2, we can use it as an instrument for y_2 in equation (9.42). Under this reasoning, we would have enough instruments—z_1, z_2, z_1^2—to identify equation (9.42). In fact, any number of functions of z_1 and z_2 can be added to the instrument list.

The fact that this argument is faulty is fortunate because our identification analysis in Section 9.2.2 says that equation (9.42) is not identified. In this example it is clear that z_1^2 cannot appear in the reduced form for y_2 because z_1^2 appears nowhere in the system. Technically, because $E(y_2 \mid \mathbf{z})$ is linear in z_1 and z_2 under assumption (9.44), the linear projection of y_2 onto (z_1, z_2, z_1^2) does not depend on z_1^2:

$$L(y_2 \mid z_1, z_2, z_1^2) = L(y_2 \mid z_1, z_2) = \pi_{21} z_1 + \pi_{22} z_2 \tag{9.45}$$

In other words, there is no *partial* correlation between y_2 and z_1^2 once z_1 and z_2 are included in the projection.

The zero conditional mean assumptions (9.41) can have some relevance for choosing an efficient estimator, although not always. If assumption (9.41) holds and $\mathrm{Var}(\mathbf{u} \mid \mathbf{z}) = \mathrm{Var}(\mathbf{u}) = \mathbf{\Sigma}$, 3SLS using instruments $\mathbf{z}$ for each equation is the asymptotically efficient estimator that uses the orthogonality conditions in assumption (9.41); this conclusion follows from Theorem 8.5. In other words, if $\mathrm{Var}(\mathbf{u} \mid \mathbf{z})$ is constant, it does not help to expand the instrument list beyond the functions of the exogenous variables actually appearing in the system.

However, if assumption (9.41) holds but $\mathrm{Var}(\mathbf{u} \mid \mathbf{z})$ is not constant, we can do better (asymptotically) than 3SLS. If $\mathbf{h}(\mathbf{z})$ is some additional functions of the exogenous variables, the minimum chi-square estimator using $[\mathbf{z}, \mathbf{h}(\mathbf{z})]$ as instruments in each equation is, generally, more efficient than 3SLS or minimum chi-square using only $\mathbf{z}$ as IVs. This result was discovered independently by Hansen (1982) and White (1982b), and it follows from the discussion in Section 8.6. Expanding the IV list to arbitrary functions of $\mathbf{z}$ and applying full GMM is not used very much in practice: it is usually not clear how to choose $\mathbf{h}(\mathbf{z})$, and, if we use too many additional instruments, the finite sample properties of the GMM estimator can be poor, as we discussed in Section 8.6.

For SEMs linear in the parameters but nonlinear in endogenous variables (in a sense to be made precise), adding nonlinear functions of the exogenous variables to the instruments not only is desirable, but is often needed to achieve identification. We turn to this topic next.

9.5 SEMs Nonlinear in Endogenous Variables

We now study models that are nonlinear in some endogenous variables. While the general estimation methods we have covered are still applicable, identification and choice of instruments require special attention.

9.5.1 Identification

The issues that arise in identifying models nonlinear in endogenous variables are most easily illustrated with a simple example. Suppose that supply and demand are given by

$$\log(q) = \gamma_{12} \log(p) + \gamma_{13}[\log(p)]^2 + \delta_{11} z_1 + u_1 \tag{9.46}$$

$$\log(q) = \gamma_{22} \log(p) + \delta_{22} z_2 + u_2 \tag{9.47}$$

$$\mathrm{E}(u_1 \mid \mathbf{z}) = \mathrm{E}(u_2 \mid \mathbf{z}) = 0 \tag{9.48}$$

where the first equation is the supply equation, the second equation is the demand equation, and the equilibrium condition that supply equals demand has been imposed. For simplicity, we do not include an intercept in either equation, but no important conclusions hinge on this omission. The exogenous variable z_1 shifts the supply function but not the demand function; z_2 shifts the demand function but not the supply function. The vector of exogenous variables appearing somewhere in the system is $\mathbf{z} = (z_1, z_2)$.

It is important to understand why equations (9.46) and (9.47) constitute a "nonlinear" system. This system is still linear in *parameters*, which is important because it means that the IV procedures we have learned up to this point are still applicable. Further, it is *not* the presence of the logarithmic transformations of q and p that makes the system nonlinear. In fact, if we set $\gamma_{13} = 0$, then the model *is* linear for the purposes of identification and estimation: defining $y_1 \equiv \log(q)$ and $y_2 \equiv \log(p)$, we can write equations (9.46) and (9.47) as a standard two-equation system.

When we include $[\log(p)]^2$ we have the model

$$y_1 = \gamma_{12} y_2 + \gamma_{13} y_2^2 + \delta_{11} z_1 + u_1 \tag{9.49}$$

$$y_1 = \gamma_{22} y_2 + \delta_{22} z_2 + u_2 \tag{9.50}$$

With this system there is no way to define *two* endogenous variables such that the system is a two-equation system in two endogenous variables. The presence of y_2^2 in equation (9.49) makes this model different from those we have studied up until now. We say that this is a system **nonlinear in endogenous variables**. What this statement really means is that, while the system is still linear in parameters, identification needs to be treated differently.

If we used equations (9.49) and (9.50) to obtain y_2 as a function of the z_1, z_2, u_1, u_2, and the parameters, the result would not be linear in $\mathbf{z}$ and $\mathbf{u}$. In this particular case we can find the solution for y_2 using the quadratic formula (assuming a real solution exists). However, $\mathrm{E}(y_2 \mid \mathbf{z})$ would not be linear in $\mathbf{z}$ unless $\gamma_{13} = 0$, and $\mathrm{E}(y_2^2 \mid \mathbf{z})$ would not be linear in $\mathbf{z}$ regardless of the value of γ_{13}. These observations have important implications for identification of equation (9.49) and for choosing instruments.

Before considering equations (9.49) and (9.50) further, consider a second example where closed form expressions for the endogenous variables in terms of the exogenous variables and structural errors do not even exist. Suppose that a system describing crime rates in terms of law enforcement spending is

$$crime = \gamma_{12}\log(spending) + \mathbf{z}_{(1)}\boldsymbol{\delta}_{(1)} + u_1 \tag{9.51}$$

$$spending = \gamma_{21}crime + \gamma_{22}crime^2 + \mathbf{z}_{(2)}\boldsymbol{\delta}_{(2)} + u_2 \tag{9.52}$$

where the errors have zero mean given $\mathbf{z}$. Here, we cannot solve for either *crime* or *spending* (or any other transformation of them) in terms of $\mathbf{z}$, u_1, u_2, and the parameters. And there is no way to define y_1 and y_2 to yield a linear SEM in two endogenous variables. The model is still linear in parameters, but $\mathrm{E}(crime \mid \mathbf{z})$, $\mathrm{E}[\log(spending) \mid \mathbf{z}]$, and $\mathrm{E}(spending \mid \mathbf{z})$ are not linear in $\mathbf{z}$ (nor can we find closed forms for these expectations).

One possible approach to identification in nonlinear SEMs is to ignore the fact that the same endogenous variables show up differently in different equations. In the supply and demand example, define $y_3 \equiv y_2^2$ and rewrite equation (9.49) as

$$y_1 = \gamma_{12}y_2 + \gamma_{13}y_3 + \delta_{11}z_1 + u_1 \tag{9.53}$$

Or, in equations (9.51) and (9.52) define $y_1 = crime$, $y_2 = spending$, $y_3 = \log(spending)$, and $y_4 = crime^2$, and write

$$y_1 = \gamma_{12}y_3 + \mathbf{z}_{(1)}\boldsymbol{\delta}_{(1)} + u_1 \tag{9.54}$$

$$y_2 = \gamma_{21}y_1 + \gamma_{22}y_4 + \mathbf{z}_{(2)}\boldsymbol{\delta}_{(2)} + u_2 \tag{9.55}$$

Defining nonlinear functions of endogenous variables as new endogenous variables turns out to work fairly generally, *provided* we apply the rank and order conditions properly. The key question is, What kinds of equations do we add to the system for the newly defined endogenous variables?

If we add linear projections of the newly defined endogenous variables in terms of the *original* exogenous variables appearing somewhere in the system—that is, the linear projection onto $\mathbf{z}$—then we are being much too restrictive. For example, suppose to equations (9.53) and (9.50) we add the linear equation

$$y_3 = \pi_{31}z_1 + \pi_{32}z_2 + v_3 \tag{9.56}$$

where, by definition, $\mathrm{E}(z_1 v_3) = \mathrm{E}(z_2 v_3) = 0$. With equation (9.56) to round out the system, the order condition for identification of equation (9.53) clearly fails: we have two endogenous variables in equation (9.53) but only one excluded exogenous variable, z_2.

The conclusion that equation (9.53) is not identified is too pessimistic. There are many other possible instruments available for y_2^2. Because $\mathrm{E}(y_2^2 \mid \mathbf{z})$ is not linear in z_1 and z_2 (even if $\gamma_{13} = 0$), other functions of z_1 and z_2 will appear in a linear projection involving y_2^2 as the dependent variable. To see what the most useful of these are likely to be, suppose that the structural system actually is linear, so that $\gamma_{13} = 0$. Then $y_2 = \pi_{21}z_1 + \pi_{22}z_2 + v_2$, where v_2 is a linear combination of u_1 and u_2. Squaring this reduced form and using $\mathrm{E}(v_2 \mid \mathbf{z}) = 0$ gives

$$\mathrm{E}(y_2^2 \mid \mathbf{z}) = \pi_{21}^2 z_1^2 + \pi_{22}^2 z_2^2 + 2\pi_{21}\pi_{22}z_1 z_2 + \mathrm{E}(v_2^2 \mid \mathbf{z}) \tag{9.57}$$

If $\mathrm{E}(v_2^2 \mid \mathbf{z})$ is constant, an assumption that holds under homoskedasticity of the structural errors, then equation (9.57) shows that y_2^2 is correlated with z_1^2, z_2^2, and $z_1 z_2$, which makes these functions natural instruments for y_2^2. The only case where no functions of $\mathbf{z}$ are correlated with y_2^2 occurs when both π_{21} and π_{22} equal zero, in which case the linear version of equation (9.49) (with $\gamma_{13} = 0$) is also unidentified.

Because we derived equation (9.57) under the restrictive assumptions $\gamma_{13} = 0$ and homoskedasticity of v_2, we would not want our linear projection for y_2^2 to omit the exogenous variables that originally appear in the system. In practice, we would augment equations (9.53) and (9.50) with the linear projection

$$y_3 = \pi_{31}z_1 + \pi_{32}z_2 + \pi_{33}z_1^2 + \pi_{34}z_2^2 + \pi_{35}z_1 z_2 + v_3 \tag{9.58}$$

where v_3 is, by definition, uncorrelated with z_1, z_2, z_1^2, z_2^2, and $z_1 z_2$. The system (9.53), (9.50), and (9.58) can now be studied using the usual rank condition.

Adding equation (9.58) to the original system and then studying the rank condition of the first two equations is equivalent to studying the rank condition in the smaller system (9.53) and (9.50). What we mean by this statement is that we do not explicitly add an equation for $y_3 = y_2^2$, but we *do* include y_3 in equation (9.53). Therefore, when applying the rank condition to equation (9.53), we use $G = 2$ (not $G = 3$). The reason this approach is the same as studying the rank condition in the three-equation system (9.53), (9.50), and (9.58) is that adding the third equation increases the rank of $\mathbf{R}_1\mathbf{B}$ by one whenever at least one additional nonlinear function of $\mathbf{z}$ appears in equation (9.58). (The functions z_1^2, z_2^2, and $z_1 z_2$ appear nowhere else in the system.)

As a general approach to identification in models where the nonlinear functions of the endogenous variables depend only on a single endogenous variable—such as the two examples that we have already covered—Fisher (1965) argues that the following method is sufficient for identification:

1. Relabel the nonredundant functions of the endogenous variables to be new endogenous variables, as in equation (9.53) or (9.54) and equation (9.55).

2. Apply the rank condition to the original system *without* increasing the number of equations. If the equation of interest satisfies the rank condition, then it is identified.

The proof that this method works is complicated, and it requires more assumptions than we have made (such as $\mathbf{u}$ being *independent* of $\mathbf{z}$). Intuitively, we can expect each additional nonlinear function of the endogenous variables to have a linear projection that depends on new functions of the exogenous variables. Each time we add another function of an endogenous variable, it effectively comes with its own instruments.

Fisher's method can be expected to work in all but the most pathological cases. One case where it does not work is if $\mathrm{E}(v_2^2 \mid \mathbf{z})$ in equation (9.57) is heteroskedastic in such a way as to cancel out the squares and cross product terms in z_1 and z_2; then $\mathrm{E}(y_2^2 \mid \mathbf{z})$ would be constant. Such unfortunate coincidences are not practically important.

It is tempting to think that Fisher's rank condition is also necessary for identification, but this is not the case. To see why, consider the two-equation system

$$y_1 = \gamma_{12} y_2 + \gamma_{13} y_2^2 + \delta_{11} z_1 + \delta_{12} z_2 + u_1 \tag{9.59}$$

$$y_2 = \gamma_{21} y_1 + \delta_{21} z_1 + u_2 \tag{9.60}$$

The first equation cleary fails the modified rank condition because it fails the order condition: there are no restrictions on the first equation except the normalization restriction. However, if $\gamma_{13} \neq 0$ and $\gamma_{21} \neq 0$, then $\mathrm{E}(y_2 \mid \mathbf{z})$ is a nonlinear function of $\mathbf{z}$ (which we cannot obtain in closed form). The result is that functions such as z_1^2, z_2^2, and $z_1 z_2$ (and others) will appear in the linear projections of y_2 and y_2^2 even after z_1 and z_2 have been included, and these can then be used as instruments for y_2 and y_2^2. But if $\gamma_{13} = 0$, the first equation cannot be identified by adding nonlinear functions of z_1 and z_2 to the instrument list: the linear projection of y_2 on z_1, z_2, and any function of (z_1, z_2) will only depend on z_1 and z_2.

Equation (9.59) is an example of a **poorly identified model** because, when it is identified, it is **identified due to a nonlinearity** ($\gamma_{13} \neq 0$ in this case). Such identification is especially tenuous because the hypothesis H_0: $\gamma_{13} = 0$ cannot be tested by estimating the structural equation (since the structural equation is not identified when H_0 holds).

There are other models where identification can be verified using reasoning similar to that used in the labor supply example. Models with interactions between exogenous variables and endogenous variables can be shown to be identified when the model without the interactions is identified (see Example 6.2 and Problem 9.6). Models with interactions among endogenous variables are also fairly easy to handle. Generally, it is good practice to check whether the most general *linear* version of the model would be identified. If it is, then the nonlinear version of the model is probably identified. We saw this result in equation (9.46): if this equation is identified when $\gamma_{13} = 0$, then it is identified for any value of γ_{13}. If the most general linear version of a nonlinear model is not identified, we should be very wary about proceeding, since identification hinges on the presence of nonlinearities that we usually will not be able to test.

9.5.2 Estimation

In practice, it is difficult to know which additional functions we should add to the instrument list for nonlinear SEMs. Naturally, we must always include the exogenous variables appearing somewhere in the system instruments in every equation. After that, the choice is somewhat arbitrary, although the functional forms appearing in the structural equations can be helpful.

A general approach is to always use some squares and cross products of the exogenous variables appearing somewhere in the system. If something like $exper^2$ appears in the system, additional terms such as $exper^3$ and $exper^4$ would be added to the instrument list.

Once we decide on a set of instruments, any equation in a nonlinear SEM can be estimated by 2SLS. Because each equation satisfies the assumptions of single-equation analysis, we can use everything we have learned up to now for inference and specification testing for 2SLS. A system method can also be used, where linear projections for the functions of endogenous variables are explicitly added to the system. Then, all exogenous variables included in these linear projections can be used as the instruments for every equation. The minimum chi-square estimator is generally more appropriate than 3SLS because the homoskedasticity assumption will rarely be satisfied in the linear projections.

It is important to apply the instrumental variables procedures directly to the structural equation or equations. In other words, we should directly use the formulas for 2SLS, 3SLS, or GMM. Trying to mimic 2SLS or 3SLS by substituting fitted values for some of the endogenous variables inside the nonlinear functions is usually a mistake: neither the conditional expectation nor the linear projection operator passes through nonlinear functions, and so such attempts rarely produce consistent estimators in nonlinear systems.

Example 9.6 (Nonlinear Labor Supply Function): We add $[\log(wage)]^2$ to the labor supply function in Example 9.5:

$$
\begin{aligned}
hours = {} & \gamma_{12} \log(wage) + \gamma_{13}[\log(wage)]^2 + \delta_{10} + \delta_{11} educ + \delta_{12} age \\
& + \delta_{13} kidslt6 + \delta_{14} kidsge6 + \delta_{15} nwifeinc + u_1 \qquad (9.61)
\end{aligned}
$$

$$\log(wage) = \delta_{20} + \delta_{21} educ + \delta_{22} exper + \delta_{23} exper^2 + u_2 \qquad (9.62)$$

where we have dropped *hours* from the wage offer function because it was insignificant in Example 9.5. The natural assumptions in this system are $\mathrm{E}(u_1 \mid \mathbf{z}) = \mathrm{E}(u_2 \mid \mathbf{z}) = 0$, where $\mathbf{z}$ contains all variables other than *hours* and $\log(wage)$.

There are many possibilities as additional instruments for $[\log(wage)]^2$. Here, we add three quadratic terms to the list—age^2, $educ^2$, and $nwifeinc^2$—and we estimate equation (9.61) by 2SLS. We obtain $\hat{\gamma}_{12} = 1{,}873.62$ (se $= 635.99$) and $\hat{\gamma}_{13} = -437.29$ (se $= 350.08$). The t statistic on $[\log(wage)]^2$ is about -1.25, so we would be justified in dropping it from the labor supply function. Regressing the 2SLS residuals $\hat{u}_1$ on all variables used as instruments in the supply equation gives R-squared $= .0061$, and so the N-R-squared statistic is 2.61. With a χ^2_3 distribution this gives p-value $= .456$. Thus, we fail to reject the overidentifying restrictions.

In the previous example we may be tempted to estimate the labor supply function using a two-step procedure that appears to mimic 2SLS:

1. Regress $\log(wage)$ on all exogenous variables appearing in the system and obtain the predicted values. For emphasis, call these $\hat{y}_2$.
2. Estimate the labor supply function from the OLS regression *hours* on 1, $\hat{y}_2$, $(\hat{y}_2)^2$, $educ, \ldots, nwifeinc$.

This two-step procedure is *not* the same as estimating equation (9.61) by 2SLS, and, except in special circumstances, it does *not* produce consistent estimators of the structural parameters. The regression in step 2 is an example of what is sometimes called a **forbidden regression**, a phrase that describes replacing a nonlinear function of an endogenous explanatory variable with the same nonlinear function of fitted values from a first-stage estimation. In plugging fitted values into equation (9.61), our mistake is in thinking that the linear projection of the square is the square of the linear projection. What the 2SLS estimator does in the first stage is project each of y_2 and y_2^2 onto the original exogenous variables and the additional nonlinear functions of these that we have chosen. The fitted values from the reduced form regression for y_2^2, say $\hat{y}_3$, are not the same as the squared fitted values from the reduced form regression for y_2, $(\hat{y}_2)^2$. This distinction is the difference between a consistent estimator and an inconsistent estimator.

If we apply the forbidden regression to equation (9.61), some of the estimates are very different from the 2SLS estimates. For example, the coefficient on *educ*, when equation (9.61) is properly estimated by 2SLS, is about -87.85 with a t statistic of -1.32. The forbidden regression gives a coefficient on *educ* of about -176.68 with a t statistic of -5.36. Unfortunately, the t statistic from the forbidden regression is generally invalid, even asymptotically. (The forbidden regression will produce consistent estimators in the special case $\gamma_{13} = 0$, if $\mathrm{E}(u_1 \mid \mathbf{z}) = 0$; see Problem 9.12.)

Many more functions of the exogenous variables could be added to the instrument list in estimating the labor supply function. From Chapter 8, we know that efficiency of GMM never falls by adding more nonlinear functions of the exogenous variables to the instrument list (even under the homoskedasticity assumption). This statement is true whether we use a single-equation or system method. Unfortunately, the fact that we do no worse asymptotically by adding instruments is of limited practical help, since we do not want to use too many instruments for a given data set. In Example 9.6, rather than using a long list of additional nonlinear functions, we might use $(\hat{y}_2)^2$ as a single IV for y_2^2. (This method is not the same as the forbidden regression!) If it happens that $\gamma_{13} = 0$ and the structural errors are homoskedastic, this would be the optimal IV. (See Problem 9.12.)

A general system linear in parameters can be written as

$$\begin{aligned} y_1 &= \mathbf{q}_1(\mathbf{y}, \mathbf{z})\boldsymbol{\beta}_1 + u_1 \\ &\vdots \\ y_G &= \mathbf{q}_G(\mathbf{y}, \mathbf{z})\boldsymbol{\beta}_G + u_G \end{aligned} \tag{9.63}$$

where $\mathrm{E}(u_g \mid \mathbf{z}) = 0$, $g = 1, 2, \ldots, G$. Among other things this system allows for complicated interactions among endogenous and exogenous variables. We will not give a general analysis of such systems because identification and choice of instruments are too abstract to be very useful. Either single-equation or system methods can be used for estimation.

9.6 Different Instruments for Different Equations

There are general classes of SEMs where the same instruments cannot be used for every equation. We already encountered one such example, the fully recursive system. Another general class of models is SEMs where, in addition to simultaneous determination of some variables, some equations contain variables that are endogenous as a result of omitted variables or measurement error.

As an example, reconsider the labor supply and wage offer equations (9.28) and (9.62), respectively. On the one hand, in the supply function it is not unreasonable to assume that variables other than log(*wage*) are uncorrelated with u_1. On the other hand, ability is a variable omitted from the log(*wage*) equation, and so *educ* might be correlated with u_2. This is an omitted variable, not a simultaneity, issue, but the statistical problem is the same: correlation between the error and an explanatory variable.

Equation (9.28) is still identified as it was before, because *educ* is exogenous in equation (9.28). What about equation (9.62)? It satisfies the order condition because we have excluded four exogenous variables from equation (9.62): *age*, *kidslt6*, *kidsge6*, and *nwifeinc*. How can we analyze the rank condition for this equation? We need to add to the system the linear projection of *educ* on all exogenous variables:

$$educ = \delta_{30} + \delta_{31} exper + \delta_{32} exper^2 + \delta_{33} age + \delta_{34} kidslt6 + \delta_{35} kidsge6 + \delta_{36} nwifeinc + u_3 \tag{9.64}$$

Provided the variables other than *exper* and $exper^2$ are sufficiently partially correlated with *educ*, the $\log(wage)$ equation is identified. However, the 2SLS estimators might be poorly behaved if the instruments are not very good. If possible, we would add other exogenous factors to equation (9.64) that are partially correlated with *educ*, such as mother's and father's education. In a system procedure, because we have assumed that *educ* is uncorrelated with u_1, *educ* can, and should, be included in the list of instruments for estimating equation (9.28).

This example shows that having different instruments for different equations changes nothing for single-equation analysis: we simply determine the valid list of instruments for the endogenous variables in the equation of interest and then estimate the equations separately by 2SLS. Instruments may be required to deal with simultaneity, omitted variables, or measurement error, in any combination.

Estimation is more complicated for system methods. First, if 3SLS is to be used, then the GMM 3SLS version must be used to produce consistent estimators of any equation; the more traditional 3SLS estimator discussed in Section 8.3.5 is generally valid only when all instruments are uncorrelated with all errors. When we have different instruments for different equations, the instrument matrix has the form in equation (8.15).

There is a more subtle issue that arises in system analysis with different instruments for different equations. While it is still popular to use 3SLS methods for such problems, it turns out that the key assumption that makes 3SLS the efficient GMM estimator, Assumption SIV.5, is often violated. In such cases the GMM estimator with general weighting matrix enhances asymptotic efficiency and simplifies inference.

As a simple example, consider a two-equation system

$$y_1 = \delta_{10} + \gamma_{12} y_2 + \delta_{11} z_1 + u_1 \tag{9.65}$$

$$y_2 = \delta_{20} + \gamma_{21} y_1 + \delta_{22} z_2 + \delta_{23} z_3 + u_2 \tag{9.66}$$

where (u_1, u_2) has mean zero and variance matrix $\mathbf{\Sigma}$. Suppose that z_1, z_2, and z_3 are uncorrelated with u_2 but we can only assume that z_1 and z_3 are uncorrelated with u_1.

In other words, z_2 is not exogenous in equation (9.65). Each equation is still identified by the order condition, and we just assume that the rank conditions also hold. The instruments for equation (9.65) are $(1, z_1, z_3)$, and the instruments for equation (9.66) are $(1, z_1, z_2, z_3)$. Write these as $\mathbf{z}_1 \equiv (1, z_1, z_3)$ and $\mathbf{z}_2 \equiv (1, z_1, z_2, z_3)$. Assumption SIV.5 requires the following three conditions:

$$\mathrm{E}(u_1^2 \mathbf{z}_1' \mathbf{z}_1) = \sigma_1^2 \mathrm{E}(\mathbf{z}_1' \mathbf{z}_1) \tag{9.67}$$

$$\mathrm{E}(u_2^2 \mathbf{z}_2' \mathbf{z}_2) = \sigma_2^2 \mathrm{E}(\mathbf{z}_2' \mathbf{z}_2) \tag{9.68}$$

$$\mathrm{E}(u_1 u_2 \mathbf{z}_1' \mathbf{z}_2) = \sigma_{12} \mathrm{E}(\mathbf{z}_1' \mathbf{z}_2) \tag{9.69}$$

The first two conditions hold if $\mathrm{E}(u_1 \mid \mathbf{z}_1) = \mathrm{E}(u_2 \mid \mathbf{z}_2) = 0$ and $\mathrm{Var}(u_1 \mid \mathbf{z}_1) = \sigma_1^2$, $\mathrm{Var}(u_2 \mid \mathbf{z}_2) = \sigma_2^2$. These are standard zero conditional mean and homoskedasticity assumptions. The potential problem comes with condition (9.69). Since u_1 is correlated with one of the elements in $\mathbf{z}_2$, we can hardly just assume condition (9.69). Generally, there is no conditioning argument that implies condition (9.69). One case where condition (9.69) holds is if $\mathrm{E}(u_2 \mid u_1, z_1, z_2, z_3) = 0$, which implies that u_2 and u_1 are uncorrelated. The left-hand side of condition (9.69) is also easily shown to equal zero. But 3SLS with $\sigma_{12} = 0$ imposed is just 2SLS equation by equation. If u_1 and u_2 are correlated, we should not expect condition (9.69) to hold, and therefore the general minimum chi-square estimator should be used for estimation and inference.

Wooldridge (1996) provides a general discussion and contains other examples of cases in which Assumption SIV.5 can and cannot be expected to hold. Whenever a system contains linear projections for nonlinear functions of endogenous variables, we should expect Assumption SIV.5 to fail.

Problems

9.1. Discuss whether each example satisfies the autonomy requirement for true simultaneous equations analysis. The specification of y_1 and y_2 means that each is to be written as a function of the other in a two-equation system.

a. For an employee, $y_1 =$ hourly wage, $y_2 =$ hourly fringe benefits.

b. At the city level, $y_1 =$ per capita crime rate, $y_2 =$ per capita law enforcement expenditures.

c. For a firm operating in a developing country, $y_1 =$ firm research and development expenditures, $y_2 =$ firm foreign technology purchases.

d. For an individual, $y_1 =$ hourly wage, $y_2 =$ alcohol consumption.