

Chapter 4

Linear Structural-Equation Models

from

Linear Statistical Models and Related Methods With Applications to Social Research

by

John Fox

1984

Copyright © by John Fox

4

Linear Structural-Equation Models

Structural-equation models enable the researcher to examine simultaneous relationships among a number of variables, some of which may exert mutual influence on each other. These models are therefore differentiated from the single-equation linear models discussed in the first three chapters of this text, which treat the relationship of one dependent variable to one or more of its causes. Structural-equation models likewise differ from direct multivariate extensions of the general linear model which, although they take into account the correlation of several dependent variables, treat these variables in parallel rather than distinguishing causal relations among them (see, for example, Morrison, 1976: Chapter 5).

In the social sciences, structural-equation models are frequently termed "causal models," and although this terminology accurately reflects the purpose of the models, it is potentially misleading for two quite different reasons. On the one hand, most data analysis in the social sciences seeks to discover causal relations: Few studies are simply correlational or predictive in their purpose. On the other hand, structural-equation models in no way avoid the pitfalls of drawing causal inferences from observational data. The term "causal model," then, at once promises too much and is nonspecific.

Yet, one of the great virtues of structural-equation models is the light they shed on the process of causal interpretation of correlational data, making explicit the assumptions underlying causal inference. These insights apply not only to the formal application of structural-equation models, but also to other methods of data analysis.

This chapter begins with a consideration of the form and specification of structural-equation models. Section 4.2 develops the method of instrumental-

variables estimation, which provides us with a tool for analyzing and estimating structural-equation models. Having specified a structural-equation model, it is necessary to determine whether the model is estimable, an issue termed the identification problem. In Section 4.3 we show how the instrumental-variables method and other approaches lead to a solution of the identification problem, after which, in Section 4.4, we describe several estimation methods applicable to structural-equation models. In Section 4.5 we explain how an estimated structural-equation model may be used for causal interpretation of statistical relationships, and we take an opportunity to draw general methodological lessons for causal inference from this discussion.

Variables in the social sciences are frequently measured with error; likewise, there is often a less-than-perfect relation between theoretical constructs and their measured indicators. There has been a consequent interest to incorporate measurement errors and multiple indicators in structural-equation models. In Section 4.6, after considering some simple examples of models with measurement errors, we present a very general model for variables measured with error. The final section of the chapter takes up the evaluation of structural-equation models that have been fit to data.

Outside of economics, the majority of applications of structural-equation models have appeared in the literature on social stratification. These applications are reflected in the illustrations and exercises of this chapter.

4.1. SPECIFICATION OF STRUCTURAL-EQUATION MODELS

This section begins by distinguishing the different categories of variables that enter into a structural-equation model, and develops graphic and equation representations of the model, introducing notational conventions along the way. We discuss the assumptions underlying the model, and define two important varieties of structural-equation models, termed recursive and block-recursive models. The section concludes by defining what is called the reduced form of a structural-equation model.

Structural-equation models include three broad classes of variables: endogenous variables, exogenous variables, and disturbances. *Endogenous variables*, as their name implies, are determined within the model, and may be influenced by other endogenous variables, by exogenous variables, and by disturbances. *Exogenous variables*, in contrast, are treated as "givens": They may appear as causes in the model, but not as effects. *Disturbance variables*, sometimes termed *errors* or *errors in equations*, represent most importantly the aggregated omitted causes of the endogenous variables, and, thus, play a role similar to that of the error variable in the general linear model. Disturbance variables are taken to be independent of the exogenous variables in the model.

4.1.1. Path Diagrams

One useful way of representing the structural relations of a model is in the form of a causal graph or *path diagram*. Consider, for example, the model

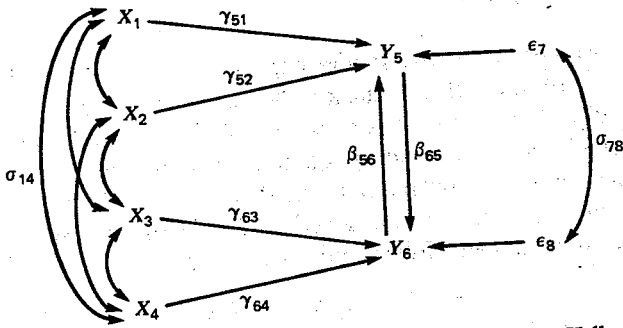


FIGURE 4.1. Nonrecursive structural-equation model for the Duncan, Haller, and Portes peer-influences data. X_1 , respondent's intelligence; X_2 , respondent's family SES; X_3 , best friend's family SES; X_4 , best friend's intelligence; Y_5 , respondent's occupational aspiration; Y_6 , best friend's occupational aspiration. (Source: see Table 4.2).

shown in Figure 4.1, from work done by Duncan, Haller, and Portes (1968) on the occupational aspirations of high-school boys. The exogenous variables in the model are represented by X 's, the endogenous variables by Y 's, and the disturbances by ϵ 's. The directed (i.e., one-way) arrows in the model indicate the direct effect of one variable on another: For example, each boy's intelligence is specified to affect directly his own aspirations, but not those of the other boy. The double-headed arrows indicate statistical relationships that are not given causal interpretation. Thus, the model, and structural-equation models in general, permit the exogenous variables to be correlated with one another. The disturbances, similarly, are not assumed to be uncorrelated. In this model, then, the aggregated omitted causes of the respondent's aspirations may be correlated with the omitted causes of his best friend's aspirations—as appears substantively sensible. Note, furthermore, that the lack of correlation¹ between exogenous variables and disturbances is reflected in the omission of double arrows linking variables in these two classes.

Each directed arrow in the path diagram is labeled with a symbol representing a *structural coefficient* of the model. As we shall see in Section 4.1.2, structural coefficients are simply regression coefficients interpreted as direct effects. γ 's are used to represent the effects of exogenous variables on endogenous variables, while β 's give the effects of endogenous variables on each other. The two subscripts of each structural parameter specify respectively the index of the effect and of its cause. γ_{51} , therefore, is the direct effect of X_1 on Y_5 , and β_{56} is the effect of Y_6 on Y_5 . The double-headed arrows are labeled with σ 's, standing for the covariances of the variables attached by the arrows. For notational convenience, each variable in the model has been assigned a unique index.

¹Throughout this chapter, we shall employ the assumptions that the disturbances and exogenous variables are independent, uncorrelated, or asymptotically uncorrelated interchangeably, according to convenience.

4.1.2. Structural Equations

The structural equations of the model express the endogenous variables as linear functions² of their direct causes. There is, therefore, in general one structural equation for each endogenous variable in the model. Although the model may initially be formulated in equation form, it is also a simple matter to read structural equations from a path diagram. A natural way of writing the structural equations of a model is as a series of (related) regression equations; for example, for the Duncan, Haller, and Portes model:

$$\begin{aligned} Y_5 &= \gamma_{51}X_1 + \gamma_{52}X_2 + \beta_{56}Y_6 + \varepsilon_7 \\ Y_6 &= \gamma_{63}X_3 + \gamma_{64}X_4 + \beta_{65}Y_5 + \varepsilon_8 \end{aligned} \quad (4.1)$$

To eliminate constant terms from the structural equations, we simply stipulate that each endogenous and exogenous variable be measured as deviations from its expectation.³ Frequently, in sociological applications, structural-equation models are specified for standardized variables, so that the coefficients of the model are standardized structural parameters. This is the case, for example, in Duncan, Haller, and Portes's research. So as not to proliferate notation, we shall not distinguish explicitly between the standardized and unstandardized cases.

Another representation of the model places every endogenous and exogenous variable in each structural equation. Variables that do not appear in a particular structural equation are given zero coefficients, and the dependent variable is given a coefficient of one. This form of the model, though more cumbersome than the regression format, has the virtue of showing explicitly which variables have been excluded from each structural equation, information that will be useful to us later on. So that variables may be aligned vertically, we shift all but the disturbances to the left side of the structural equations. For the illustrative model,

$$\begin{aligned} 1Y_5 - \beta_{56}Y_6 - \gamma_{51}X_1 - \gamma_{52}X_2 + 0X_3 + 0X_4 &= \varepsilon_7 \\ -\beta_{65}Y_5 + 1Y_6 + 0X_1 + 0X_2 - \gamma_{63}X_3 - \gamma_{64}X_4 &= \varepsilon_8 \end{aligned}$$

Finally, for compactness and generality, we write the structural-equation model as a matrix equation:

$$\mathbf{B} \begin{matrix} y_i \\ (q \times q) \end{matrix} + \mathbf{\Gamma} \begin{matrix} x_i \\ (q \times m) \end{matrix} = \begin{matrix} \varepsilon_i \\ (m \times 1) \end{matrix} \quad (4.2)$$

²As in the single-equation linear model, we require only that the structural equations be linear in the parameters. Essentially nonlinear structural-equation models are beyond the scope of this chapter: see Amemiya (1974, 1977) and Gallant (1977).

³So as to avoid complicating the notation in this chapter, unless otherwise noted, we do not use asterisks to indicate that variables are in mean-deviation form. To specify a constant term for a structural equation, it is merely necessary to leave variables in raw-score form and to include as a regressor a dummy exogenous variable coded one for each observation. Constants are rarely of substantive interest, however.

The vectors y_i , x_i , and ε_i contain endogenous variables, exogenous variables, and disturbances, each for the i th observation of a sample. \mathbf{B} contains the structural coefficients relating the endogenous variables to each other, while Γ contains the coefficients relating the endogenous to the exogenous variables. Each row of the parameter matrices includes the coefficients for one structural equation of the model, and we order the equations so that ones appear on the diagonal of \mathbf{B} . The matrix representation of the Duncan, Haller, and Portes model is shown in equation (4.3). As a matter of convenience, we have omitted the subscript i for observation.

$$\begin{pmatrix} 1 & -\beta_{56} \\ -\beta_{65} & 1 \end{pmatrix} \begin{pmatrix} Y_5 \\ Y_6 \end{pmatrix} + \begin{pmatrix} -\gamma_{51} & -\gamma_{52} & 0 & 0 \\ 0 & 0 & -\gamma_{63} & -\gamma_{64} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} = \begin{pmatrix} \varepsilon_7 \\ \varepsilon_8 \end{pmatrix} \quad (4.3)$$

Sometimes we shall require the structural equations for a sample of n observations:

$$\mathbf{Y} \quad \mathbf{B}' + \mathbf{X} \quad \Gamma' = \mathbf{E}$$

$(n \times q) \quad (q \times q) \quad (n \times m) \quad (m \times q) \quad (n \times q)$

Here, we have transposed the matrices of structural parameters, writing equations as columns, so that each observation comprises a row of \mathbf{Y} , \mathbf{X} , and \mathbf{E} .

4.1.3. Assumptions Underlying the Model

The assumptions underlying a structural-equation model are of two general types: first, assumptions of causal structure captured in the structural equations of the model; and second, distributional assumptions regarding the errors. Assumptions of causal structure are implicit in the specification of \mathbf{B} and Γ , certain of whose entries are prespecified to be zero, and in the choice of endogenous and exogenous variables.

We have already remarked that the exogenous variables and disturbances are defined (i.e., in an application, assumed) to be uncorrelated. It is often convenient to express this assumption in terms of a probability limit:

$$\text{plim} \frac{1}{n} \mathbf{X}'\mathbf{E} = \mathbf{0}$$

$(m \times q)$

The remaining assumptions about the distribution of the disturbances are analogous to the assumptions concerning the error in the general linear model: that the observations on each disturbance are independently and normally distributed with expectation zero and common variance. Note that although we

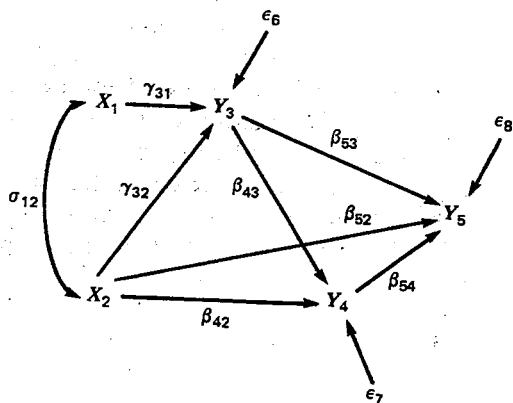


FIGURE 4.2. Recursive structural-equation model for the Blau and Duncan stratification data. X_1 , father's education; X_2 , father's occupational status; Y_3 , education; Y_4 , status of first job; Y_5 , 1962 occupational status.

assume independent observations, we in general expect *different* disturbance variables to be correlated. The joint distribution of ϵ_i is assumed to be multivariate normal with covariance matrix $\Sigma_{\epsilon\epsilon}$.

4.1.4. Recursive and Block-Recursive Models

Although structural-equation models do not in general require that different disturbance variables be independent, such assumptions may be made. In conjunction with special patterns of restrictions on the structural coefficients of the model, restrictions on disturbance covariances serve to define two important varieties of structural-equation models: *recursive* and *block-recursive* models. Models that do not satisfy the special requirements of recursive and block-recursive structures are termed *nonrecursive*. As we shall discover in Sections 4.3 and 4.4, the classification of a model has implications for its identification and estimation.

An example of a recursive structural-equation model, taken from work on stratification by Blau and Duncan (1967), is shown in Figure 4.2. This model is recursive because it meets two special conditions: (1) Different disturbance variables are specified to be uncorrelated—a characteristic reflected in the absence of bidirectional arrows linking the disturbances; and (2) the causal



FIGURE 4.3. Relations ruled out by recursive structure. (a) Reciprocal effects. (b) Causal loop.

structure of the model is unidirectional—there are no reciprocal paths or causal loops of the sort illustrated in Figure 4.3.

In the matrix representation of the model, the uncorrelated disturbances of a recursive model imply a diagonal covariance matrix of disturbances. The unidirectional causal structure implies a lower-triangular \mathbf{B} matrix, or a \mathbf{B} matrix that can be made triangular by a reordering of the endogenous variables. For the Blau and Duncan stratification model

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 \\ -\beta_{43} & 1 & 0 \\ -\beta_{53} & -\beta_{54} & 1 \end{pmatrix}$$

$$\Sigma_{ee} = \begin{pmatrix} \sigma_6^2 & 0 & 0 \\ 0 & \sigma_7^2 & 0 \\ 0 & 0 & \sigma_8^2 \end{pmatrix}$$

It should be stressed that the special requirements of a recursive model, including the stipulation of uncorrelated disturbances, must be justifiable on substantive grounds, as is the case generally for the application of statistical models. In the Blau and Duncan model, for example, we may question the independence of ε_7 and ε_8 , for Y_4 and Y_5 are likely to have common omitted causes.

It may be the case that a structural-equation model is not recursive, but that the requirements for a recursive model are met for subsets (termed *blocks*) of the endogenous variables and associated disturbances, rather than for these variables treated individually. That is, if we partition the endogenous variables and disturbances into blocks: (1) causation is unidirectional between blocks; and (2) errors are uncorrelated between blocks. Within blocks, mutual causation and correlated disturbances are permitted.

A block-recursive model is shown in Figure 4.4; this model is a modification of one specified by Duncan, Haller, and Portes (1968). Here Y_5 and Y_6 together with the associated errors ε_9 and ε_{10} comprise the first block, and Y_7 , Y_8 , ε_{11} , and ε_{12} comprise the second block. For this model,

$$\mathbf{B} = \left(\begin{array}{cc|cc} 1 & -\beta_{56} & 0 & 0 \\ -\beta_{65} & 1 & 0 & 0 \\ \hline -\beta_{75} & 0 & 1 & -\beta_{78} \\ 0 & -\beta_{86} & -\beta_{87} & 1 \end{array} \right)$$

$$\Sigma_{ee} = \left(\begin{array}{cc|cc} \sigma_{99} & \sigma_{9,10} & 0 & 0 \\ \sigma_{9,10} & \sigma_{10,10} & 0 & 0 \\ \hline 0 & 0 & \sigma_{11,11} & \sigma_{11,12} \\ 0 & 0 & \sigma_{11,12} & \sigma_{12,12} \end{array} \right)$$

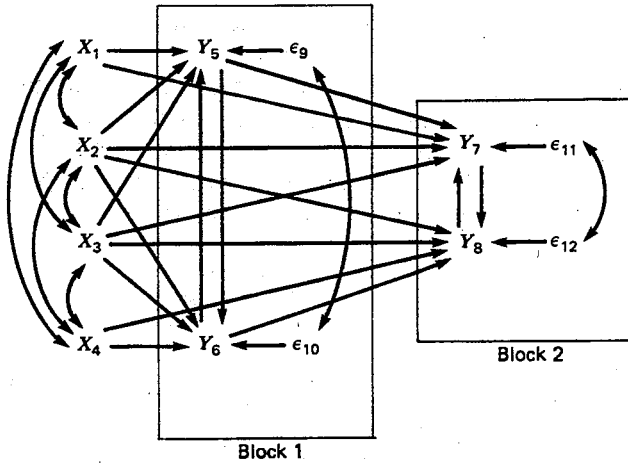


FIGURE 4.4. Block-recursive model for the Duncan, Haller, and Portes peer-influences data. X_1 , respondent's intelligence; X_2 , respondent's family SES; X_3 , best friend's family SES; X_4 , best friend's intelligence; Y_5 , respondent's occupational aspiration; Y_6 , best friend's occupational aspiration; Y_7 , respondent's educational aspiration; Y_8 , best friend's educational aspiration. (Source: see Table 4.2.)

Note that \mathbf{B} , though not triangular, is *block triangular* when partitioned according to blocks of endogenous variables, and that $\Sigma_{\epsilon\epsilon}$ is *block diagonal* when partitioned by blocks of disturbances:

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{0} \\ \mathbf{B}_{12} & \mathbf{B}_{22} \end{pmatrix}$$

$$\Sigma_{\epsilon\epsilon} = \begin{pmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{pmatrix}$$

These two characteristics, of course, may be extended to models with more than two blocks and with different numbers of endogenous variables in each block. For the peer-influences data, the specification of uncorrelated errors between blocks is questionable, for the residual causes of a boy's educational aspirations are likely also to affect his occupational aspirations.

4.1.5. The Reduced Form of the Model

Thus far, we have dealt with the structural equations of a simultaneous-equation model, equations that specify the direct causal relations among the variables in the model. The *reduced form* of a structural-equation model expresses the endogenous variables in terms of the exogenous variables and disturbances, which comprise the ultimate inputs to the system under study.

Solving the structural equations (4.2) for y is straightforward:

$$\begin{aligned}
 y &= -B^{-1}\Gamma x + B^{-1}\epsilon \\
 &= \underset{(q \times m)}{\Pi} \underset{(m \times 1)}{x} + \underset{(q \times 1)}{\delta}
 \end{aligned}
 \tag{4.4}$$

where $\Pi \equiv -B^{-1}\Gamma$ is the matrix of *reduced-form coefficients*, and $\delta \equiv B^{-1}\epsilon$ is the vector of *reduced-form errors*. Note that Π is a function of the structural parameters (B and Γ), and that δ results from a linear transformation of the structural disturbances ϵ .

In solving for y we have implicitly assumed that B is nonsingular, and we now make this condition a requirement for a well formed structural-equation model. This requirement is not problematic, however, since our method of constructing structural equations, which places ones on the main diagonal of B , virtually assures that B will be nonsingular.⁴

In the reduced form, x and δ are uncorrelated, because x contains the exogenous variables, while δ is a linear transformation of the structural disturbances ϵ . The reduced form, therefore, meets the assumptions of ordinary least-squares (*OLS*) estimation, since the independent variables in the reduced form (x) are uncorrelated with the errors (δ). We shall see later (Section 4.4) that *OLS* estimation does not in general provide consistent estimators of the structural-form parameters.

The reduced form has several uses: (1) It traces the indirect, as well as the direct impact of the exogenous variables on the endogenous variables (see Section 4.5); (2) it is useful in certain forecasting applications;⁵ and (3) it is useful in deriving a procedure for determining the estimability of a structural-equation model, a topic taken up in Section 4.3.

We gave the structural form of the Duncan, Haller, and Portes peer-influences model in equation (4.3). The reduced form of this model is

$$\begin{aligned}
 \begin{pmatrix} Y_5 \\ Y_6 \end{pmatrix} &= - \begin{pmatrix} 1 & -\beta_{56} \\ -\beta_{65} & 1 \end{pmatrix}^{-1} \begin{pmatrix} -\gamma_{51} & -\gamma_{52} & 0 & 0 \\ 0 & 0 & -\gamma_{63} & -\gamma_{64} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} \\
 &+ \begin{pmatrix} 1 & -\beta_{56} \\ -\beta_{65} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \epsilon_7 \\ \epsilon_8 \end{pmatrix}
 \end{aligned}$$

⁴The assignment of a coefficient of one to an endogenous variable in each structural equation is sometimes called a *normalization rule*. If we view a structural equation as simply specifying a relation among the endogenous and exogenous variables of the model, the normalization employed may be regarded as arbitrary. For models representable as causal diagrams, it is natural to normalize for the dependent variable in each structural equation. Some estimation methods (such as 2SLS, see Section 4.4.1) are sensitive to the normalization applied, while others (e.g., FIML, also in Section 4.4.1) are not.

⁵In economics, structural-equation models are typically applied to time series, rather than to cross-sectional data.

which, upon manipulation, yields

$$\begin{aligned}
 Y_5 &= \frac{\gamma_{51}}{1 - \beta_{56}\beta_{65}} X_1 + \frac{\gamma_{52}}{1 - \beta_{56}\beta_{65}} X_2 + \frac{\beta_{56}\gamma_{63}}{1 - \beta_{56}\beta_{65}} X_3 \\
 &\quad + \frac{\beta_{56}\gamma_{64}}{1 - \beta_{56}\beta_{65}} X_4 + \left(\frac{1}{1 - \beta_{56}\beta_{65}} \varepsilon_7 + \frac{\beta_{56}}{1 - \beta_{56}\beta_{65}} \varepsilon_8 \right) \\
 &= \pi_{51} X_1 + \pi_{52} X_2 + \pi_{53} X_3 + \pi_{54} X_4 + \delta_5 \\
 Y_6 &= \frac{\beta_{65}\gamma_{51}}{1 - \beta_{56}\beta_{65}} X_1 + \frac{\beta_{65}\gamma_{52}}{1 - \beta_{56}\beta_{65}} X_2 + \frac{\gamma_{63}}{1 - \beta_{56}\beta_{65}} X_3 \\
 &\quad + \frac{\gamma_{64}}{1 - \beta_{56}\beta_{65}} X_4 + \left(\frac{\beta_{65}}{1 - \beta_{56}\beta_{65}} \varepsilon_7 + \frac{1}{1 - \beta_{56}\beta_{65}} \varepsilon_8 \right) \\
 &= \pi_{61} X_1 + \pi_{62} X_2 + \pi_{63} X_3 + \pi_{64} X_4 + \delta_6
 \end{aligned}$$

PROBLEMS

- 4.1. For each of the path diagrams shown in Figure 4.5: (i) write out the structural equations of the model; (ii) determine whether the model is nonrecursive, recursive, or block recursive; and (iii) find the reduced form of the model.
- 4.2. The model in Figure 4.5(g) was employed by Rindfuss, Bumpass, and St. John (1980) to study the possibly reciprocal relationship between women's education and fertility. The variables in the model are:

X_1	respondent's father's occupational status
X_2	respondent's race, coded one for blacks and zero otherwise
X_3	number of respondent's siblings
X_4	farm background, coded one if the respondent grew up on a farm
X_5	regional background, coded one if the respondent grew up in the South
X_6	household composition when the respondent was 14 years old, coded zero if both parents were present in the household and one otherwise
X_7	religion, coded one if the respondent is Catholic
X_8	smoking, coded one if the respondent smoked prior to age 16
X_9	fecundity, coded one if the respondent had a miscarriage prior to the birth of her first child
Y_{10}	respondent's education at first marriage, in years
Y_{11}	age at first birth

This model was fit to data from the 1970 National Fertility Survey, for a sample of 1766 black and white American women, age 35 to 40, who had one or more children. Comment on the specification of the model, paying particular attention to its causal structure and to the distributional assumptions concerning the errors.

- 4.3. The model diagrammed in Figure 4.5(h) was specified by Berk and Berk (1978) to account for the division of household labor among the members of a family. The model was estimated for a sample of 184 households in Evanston, Illinois, an affluent suburb of Chicago. The variables

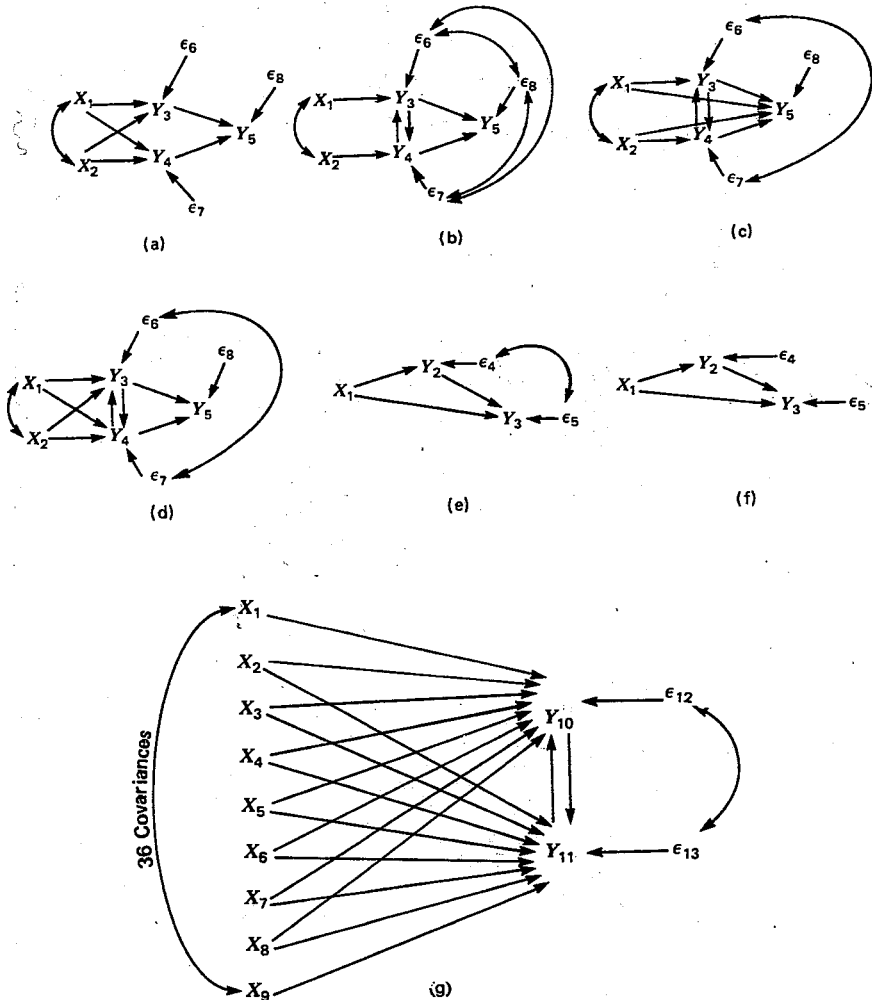


FIGURE 4.5. Path diagrams for structural-equation models. (g) Rindfuss, Bumpass, and St. John's model. (h) Berk and Berk's model. (i) Duncan, Featherman, and Duncan's model (adapted with permission from Duncan, Featherman, and Duncan, 1972). (j) Lincoln's model.

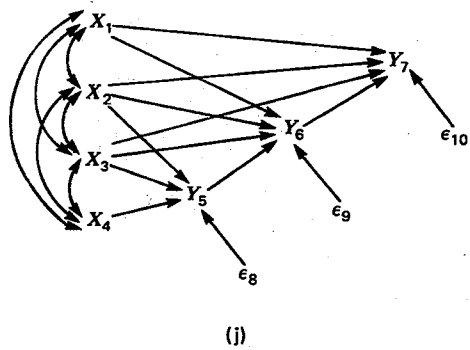
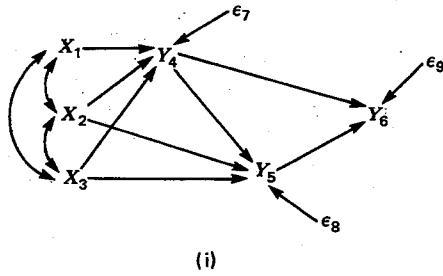
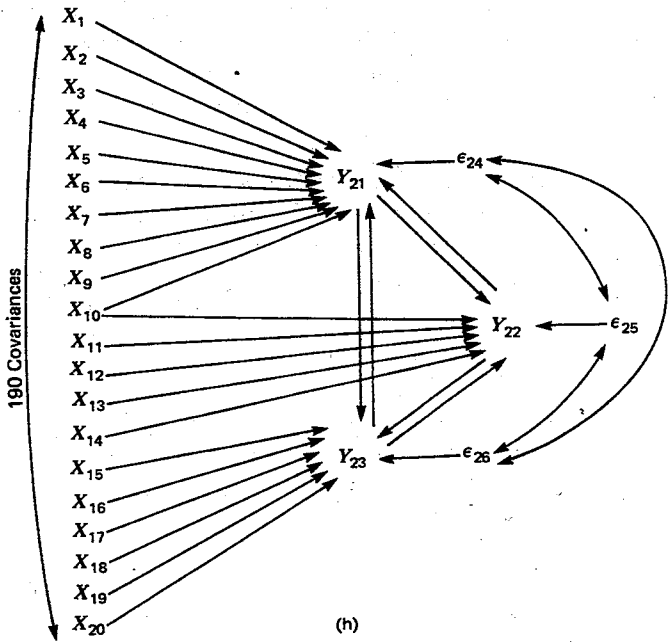


FIGURE 4.5. (Continued).

appearing in the model include:

- X_1 wife's wages, in \$1000s/month
- X_2 coded one if the wife is employed as a professional or technical worker, zero otherwise
- X_3 coded one if the wife is employed as a manager or proprietor
- X_4 coded one if there is a boy under one-year old in the home
- X_5 coded one if there is a girl under one-year old
- X_6 coded one if there is a one-year-old boy
- X_7 coded one if there is a one-year-old girl
- X_8 coded one if there is a two-year-old boy
- X_9 coded one if there is a two-year-old girl
- X_{10} year married
- X_{11} husband's monthly income
- X_{12} coded one if the husband is employed as a professional or technical worker
- X_{13} coded one if the husband is employed as a manager or proprietor
- X_{14} coded one if, in the recent past, the husband decided to do more housework
- X_{15} coded one if there is a boy 11-15
- X_{16} coded one if there is a girl 11-15
- X_{17} coded one if there is a boy 16-20
- X_{18} coded one if there is a girl 16-20
- X_{19} husband's education, measured in seven levels
- X_{20} coded one if, in the recent past, a child decided to do more housework
- Y_{21} proportion of household tasks generally done by the wife
- Y_{22} proportion of tasks generally done by the husband
- Y_{23} proportion of tasks generally done by children

Note: Y_{21} , Y_{22} , and Y_{23} are *not* constrained to sum to one, since different members of the household can regularly contribute to the same tasks. As in Problem 4.2, comment on the specification of the model.

4.4. The model in Figure 4.5(i) appears in Duncan, Featherman, and Duncan's (1972) monograph on socioeconomic background and achievement. This model, which was fit separately to each of several age groups of men, was

estimated for the same 1962 sample as employed in Blau and Duncan's (1967) study. The variables are defined as follows:

X_1	father's education
X_2	father's occupational status
X_3	number of respondent's siblings
Y_4	respondent's education
Y_5	respondent's occupational status
Y_6	respondent's income

Discuss the specification of Duncan, Featherman, and Duncan's model.

- 4.5. In a study of strike activity in metropolitan areas of the United States, Lincoln (1978) specified the structural-equation model shown in Figure 4.5(j). The model was estimated using data for 78 metropolitan areas. The variables in the model are:

X_1	the degree of concentration of union staff in the metropolitan area; this index is high when most union staff members work for a relatively small number of large unions
X_2	the degree of concentration in employment
X_3	the log of the number of employed workers
X_4	the proportion of workers who are in unionized establishments
Y_5	the number of strikes in the period 1963-1969
Y_6	the number of strikers
Y_7	the number of person-days idle due to strikes

Comment on the specification of Lincoln's model.

4.2. INSTRUMENTAL-VARIABLES ESTIMATION

After specifying a structural-equation model, it is necessary to determine whether the parameters of the model may be estimated. This issue is called the identification problem, and it will be taken up in the next section. The present section is, therefore, a necessary detour, for instrumental-variables estimation will provide us with an approach to the identification problem as well as with a means for approaching the topic of estimation (Section 4.4). Although our ultimate interest is in applying the method of instrumental variables to structural-equation models, we shall develop the method in a more general context.

Suppose, to begin, that we wish to estimate the simple-regression model

$$Y = \alpha + \beta X + \varepsilon \quad (4.5)$$

and that we make the usual assumptions that $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma_\varepsilon^2$, and X and ε are independent. (Here, X and Y are in raw-score form.) In Chapter 1, this model and these assumptions were employed in conjunction with least-squares estimation. We shall now derive the ordinary-least-squares estimator of β in an alternative manner, termed the *expectation method*. Let us express both X and Y as deviations from their expectations; that is, $Y^* \equiv Y - E(Y)$, and $X^* \equiv X - E(X)$. Then, because of the assumption that $E(\varepsilon) = 0$, the model (4.5) becomes $Y^* = \beta X^* + \varepsilon$. Multiplying this equation through by X^* and taking expectations of both sides, we get⁶

$$E(X^*Y^*) = \beta E(X^{*2}) + E(X^*\varepsilon) \quad (4.6)$$

$$\sigma_{XY} = \beta\sigma_X^2 + \sigma_{X\varepsilon}$$

Here, σ_{XY} is the population covariance of X and Y , σ_X^2 is the population variance of X , and $\sigma_{X\varepsilon}$ is the population covariance of X and ε ; this last quantity is zero due to the stipulation that X and ε are independent. Solving equation (4.6) for β gives us $\beta = \sigma_{XY}/\sigma_X^2$.

We cannot of course apply this result without knowledge of the population quantities σ_{XY} and σ_X^2 , knowledge that is generally unavailable. We can, however, estimate these parameters from sample data, employing the sample variance $S_X^2 = \sum(X_i - \bar{X})^2/(n-1)$ and the sample covariance $S_{XY} = \sum[(X_i - \bar{X})(Y_i - \bar{Y})]/(n-1)$. We know that S_X^2 and S_{XY} are consistent estimators; that is, $\text{plim } S_X^2 = \sigma_X^2$, and $\text{plim } S_{XY} = \sigma_{XY}$. The estimator $B = S_{XY}/S_X^2$ is, therefore, also consistent, for

$$\text{plim } B = \frac{\text{plim } S_{XY}}{\text{plim } S_X^2} = \frac{\sigma_{XY}}{\sigma_X^2} = \beta$$

Thus far, we have shown nothing new, because we recognize B as the usual OLS estimator of β .

Suppose, however, that we cannot assume the uncorrelation of X and ε in model (4.5), which justified the crucial elimination of $\sigma_{X\varepsilon}$ from equation (4.6), but that there is some variable Z for which it may reasonably be assumed that $\text{plim } S_{Z\varepsilon} = \sigma_{Z\varepsilon} = 0$, and that $\text{plim } S_{ZX} = \sigma_{ZX} \neq 0$. In words, Z and ε are uncorrelated in the population, but Z and X are correlated. Then, following the previous development, but multiplying through by $Z^* \equiv Z - E(Z)$ rather

⁶We assume throughout that expectations, variances, and covariances exist.

than by X^* , we obtain

$$E(Z^*Y^*) = \beta E(Z^*X^*) + E(Z^*\varepsilon)$$

$$\sigma_{ZY} = \beta\sigma_{ZX} + \sigma_{Z\varepsilon} \quad (4.7)$$

$$\beta = \frac{\sigma_{ZY}}{\sigma_{ZX}}$$

Replacing the population covariances in equation (4.7) with their consistent estimators produces a consistent estimator of β : $B = S_{ZY}/S_{ZX}$. Here, B is called an *instrumental-variable (IV) estimator*, which is generally distinct from the OLS estimator, and Z is an *instrumental variable*. Recall that the two critical requirements for an instrumental variable are uncorrelation with the error ($\sigma_{Z\varepsilon} = 0$) and nonzero correlation with the independent variable ($\sigma_{ZX} \neq 0$). OLS, then, may be thought of as a type of IV estimation, for which the instrumental variable and the independent variable are one and the same.

The method of instrumental variables may be generalized to models with several regressors, for which purpose we cast the model in matrix form:

$$Y^* = (X_1^*, X_2^*, \dots, X_k^*) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \varepsilon$$

$$= \mathbf{x}^*\boldsymbol{\beta} + \varepsilon \quad (4.8)$$

where Y^* and the X_j^* are written as deviations from their expectations, eliminating the constant term from $\boldsymbol{\beta}$. Suppose that we have available k instrumental variables in a vector \mathbf{z}^* , the entries of which are also in deviation form. We require that

$$\text{plim}_{(k \times 1)} \mathbf{s}_{Z\varepsilon} = \boldsymbol{\sigma}_{Z\varepsilon} = \mathbf{0}$$

$$\text{plim}_{(k \times k)} \mathbf{S}_{ZX} = \boldsymbol{\Sigma}_{ZX} \text{ nonsingular}$$

where $\mathbf{s}_{Z\varepsilon}$ and \mathbf{S}_{ZX} contain sample covariances, and $\boldsymbol{\sigma}_{Z\varepsilon}$ and $\boldsymbol{\Sigma}_{ZX}$ contain the corresponding population covariances. The first criterion specifies that the instrumental variables are uncorrelated with the error in the population; the second criterion requires that the instrumental variables are correlated with the independent variables and that there is not perfect collinearity. Later on, to obtain an asymptotic covariance matrix for the IV estimators, we shall also

require the existence of the population covariance matrix for the instrumental variables: $\text{plim } S_{ZZ} = \Sigma_{ZZ}$.

Proceeding as before, we premultiply both sides of equation (4.8) by z^* and take expectations:

$$E(z^*Y^*) = E(z^*X^*)\beta + E(z^*\varepsilon)$$

$$\sigma_{ZY} = \Sigma_{ZX}\beta + \sigma_{Z\varepsilon}$$

$$\beta = \Sigma_{ZX}^{-1}\sigma_{ZY}$$

Substituting sample covariances for population covariances, we then obtain

$$b = S_{ZX}^{-1}s_{ZY} = (Z^*X^*)^{-1}Z^*y^* \tag{4.9}$$

as the IV estimator of β . Because S_{ZX} and s_{ZY} are consistent estimators, so is b . In equation (4.9), Z^* , X^* , and y^* are data matrices of variables in mean-deviation form.

The asymptotic covariance matrix of the IV estimator is given by

$$\mathcal{V}(b) = \frac{\sigma_{\varepsilon}^2}{n} \Sigma_{ZX}^{-1} \Sigma_{ZZ} \Sigma_{XZ}^{-1} \tag{4.10}$$

(A relatively simple but flawed proof of equation (4.10) is given in Johnston (1972: Chapter 9); see McCallum (1973) for the correction.) Because in applications we are not in a position to know the population quantities in equation (4.10), the covariance matrix for b must be estimated. We may proceed as follows:⁷

$$e = y - Xb$$

$$S_E^2 = \frac{e'e}{n - k - 1} \tag{4.11}$$

$$\begin{aligned} \widehat{\mathcal{V}}(b) &= \frac{S_E^2}{n - 1} S_{ZX}^{-1} S_{ZZ} S_{XZ}^{-1} \\ &= S_E^2 (Z^*X^*)^{-1} Z^*Z^*(X^*Z^*)^{-1} \end{aligned}$$

For the sampling variances of b to be small, there must be large covariances between the instrumental variables and the regressors. For example, in the

⁷Since these results are asymptotic, it is also reasonable to calculate $S_E^2 = e'e/n$. Using "degrees of freedom" $n - k - 1$ rather than n in the denominator of S_E^2 produces a larger estimate of error variance and, hence, is conservative.

simple-regression model (4.5), for which $B = S_{ZY}/S_{ZX}$, the asymptotic variance of B is

$$\mathcal{V}(B) = \frac{\sigma_\varepsilon^2}{n} \left(\frac{\sigma_Z^2}{\sigma_{ZX}^2} \right) = \frac{\sigma_\varepsilon^2}{n} \left(\frac{1}{\rho_{ZX}^2 \sigma_X^2} \right)$$

where ρ_{ZX} is the population correlation of Z and X .

PROBLEMS

- 4.6. Show how the expectation method may be used to derive the OLS estimator in multiple regression. Show that OLS estimation is a type of IV estimation.
- 4.7.† Consider the multiple-regression model $\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, for n observations and k independent variables. Let \mathbf{Z}^* be an $(n \times k)$ matrix of instrumental variables, producing the IV estimator $\mathbf{b} = (\mathbf{Z}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{Z}^{*\prime}\mathbf{y}^*$. Let \mathbf{T} be any $(k \times k)$ nonsingular matrix. Show that
- $\tilde{\mathbf{Z}}^* \equiv \mathbf{Z}^*\mathbf{T}$ is also a matrix of instrumental variables; and that
 - using $\tilde{\mathbf{Z}}^*$ in place of \mathbf{Z}^* produces the same IV estimates.
 - Why is it therefore valid to conclude that what is important about a set of instrumental variables is the subspace that it spans and not the basis selected for this subspace?
 - * Figure 4.6 shows the vector geometry of the one-independent variable case, that is for the model $\mathbf{y}^* = \beta\mathbf{x}^* + \boldsymbol{\varepsilon}$. In this figure, \mathbf{x}^* is the independent-variable vector for a particular sample; $\bar{\boldsymbol{\varepsilon}} \equiv E(\boldsymbol{\varepsilon}|\mathbf{x}^*)$ (why is it nonzero?); $\bar{\mathbf{z}}^* \equiv E(\mathbf{z}^*|\mathbf{x}^*)$ (why, from the figure, is Z qualified to be an IV?); and $\bar{\mathbf{y}}^* \equiv E(\mathbf{y}^*|\mathbf{x}^*) = \beta\mathbf{x}^* + \bar{\boldsymbol{\varepsilon}}$ (note that $E(\mathbf{y}^*) \neq \bar{\mathbf{y}}^*$; why?). In drawing Figure 4.6, we take a line of sight perpendicular to the $\mathbf{x}^*, \bar{\mathbf{z}}^*$ plane; in other words, we may think of $\bar{\boldsymbol{\varepsilon}}$ and $\bar{\mathbf{y}}^*$ as orthogonal projections onto this plane. On the basis of this figure, and working with the population analogs of the estimators, explain why (i) the OLS estimator is biased, and (ii) the IV estimator is not. [Hint: Show that the triangles $(\mathbf{0}, \bar{\mathbf{y}}^*, \beta\mathbf{x}^*)$ and $(\mathbf{0}, \bar{\mathbf{x}}^*, \mathbf{x}^*)$ are similar and thus $\beta = \|\beta\mathbf{x}^*\|/\|\mathbf{x}^*\| = \|\bar{\mathbf{y}}^*\|/\|\bar{\mathbf{x}}^*\|$; find expressions for $\bar{\mathbf{y}}^*$ and $\bar{\mathbf{x}}^*$ using the fact that they are orthogonal projections onto $\bar{\mathbf{z}}^*$, and substitute these into the formula for β .] Note that in this problem we are dealing with the population analogs of the estimators. Since \mathbf{x}^* is not fixed over repeated sampling, these results expressed in terms of expectations hold only roughly in finite samples. Further information on the geometry of IV estimation may be found in Wonnacott and Wonnacott (1979: 453–455).

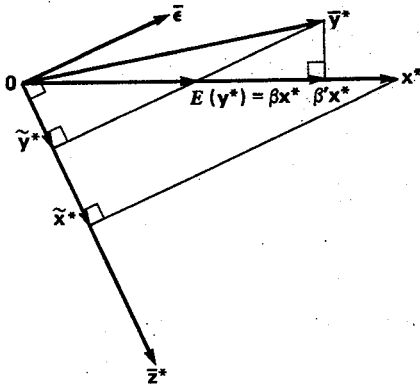


FIGURE 4.6. Vector geometry of instrumental-variable estimation. (Adapted with permission from Wonnacott and Wonnacott, 1979: 454.)

4.3. THE IDENTIFICATION PROBLEM AND ITS SOLUTION

As we have mentioned, once a structural-equation model is specified, it is necessary to determine whether the parameters of the model can be estimated. This issue of estimability is called the *identification problem*. In single-equation linear models with full-rank design matrices, the general assumptions of the model assure that its parameters may be estimated. In structural-equation models, the distributional assumptions of the model are generally insufficient to guarantee identification of its parameters; to assure identification, additional assumptions, taking the form of *a priori* restrictions on the parameters of the model, are necessary.

In general, two sorts of prior restrictions are placed on the model: (1) restrictions on structural parameters, typically specifying that certain parameters are zero; and (2) restrictions on covariances between disturbances, typically specifying that certain of these covariances are zero. We shall first consider the identification of nonrecursive models, where no restrictions are placed on disturbance covariances, deriving general rules for determining whether a model is identified. Then we shall take up the identification of recursive and block-recursive models which, as we know, specify that certain disturbance covariances are zero. Finally, we shall examine the identification status of nonrecursive models that place restrictions on covariances between disturbances.

A parameter in a structural-equation model is *identified* if it can be estimated, and *underidentified* (or *unidentified*) otherwise. If more than one estimator of the parameter can be obtained, then the parameter is *overidentified*; if just one estimator can be obtained, then the parameter is *exactly* (or *just*) *identified*. These distinctions are illustrated in Figure 4.7. The same terminology is applicable to structural equations and to the structural-equation model as a whole. Thus, a structural equation is just identified if there is one and only one way of estimating its parameters. Likewise, a model is overidentified if all its structural equations are identified, and if at least one structural

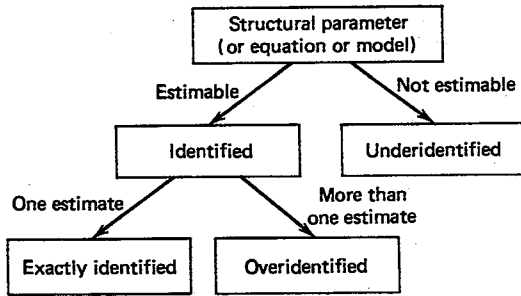


FIGURE 4.7. Identification terminology.

equation is overidentified. In practice, the identification status of a model is generally determined one structural equation at a time.

4.3.1. Identification of Nonrecursive Models

There are several approaches that can be taken to the identification problem. The simplest approach employs the method of instrumental variables, and it is with this approach that we begin. The IV approach yields the so-called order condition, which is a necessary-but-not-sufficient condition for identification. After pursuing the IV approach, we develop a method based on transformations of the structural equations and on the relationship of the structural equations to the reduced form. This method produces a necessary-and-sufficient condition for identification called the rank condition.

The Instrumental-Variables Approach Because they are independent of the disturbances, the exogenous variables of a structural-equation model provide a pool of instrumental variables for estimating the structural parameters of the model. Consider the first structural equation of the Duncan, Haller, and Portes model, given in equation (4.1) and repeated here:

$$Y_5 = \gamma_{51}X_1 + \gamma_{52}X_2 + \beta_{56}Y_6 + \varepsilon_7 \tag{4.12}$$

Multiplying equation (4.12) through by the exogenous variables, taking expectations, and substituting sample covariances for population covariances produces four IV estimating equations:

IVs	Estimating Equations
X_1	$S_{15} = C_{51}S_{11} + C_{52}S_{12} + B_{56}S_{16}$
X_2	$S_{25} = C_{51}S_{12} + C_{52}S_{22} + B_{56}S_{26}$
X_3	$S_{35} = C_{51}S_{13} + C_{52}S_{23} + B_{56}S_{36}$
X_4	$S_{45} = C_{51}S_{14} + C_{52}S_{24} + B_{56}S_{46}$

(4.13)

Here there are four estimating equations in only three unknowns (C_{51} , C_{52} , and B_{56}), since the sample covariances represent known coefficients calculable from sample data. In general, therefore, the estimating equations (4.13) will be overdetermined: For an actual sample, there will in general be no set of values c_{51} , c_{52} , and b_{56} that simultaneously satisfies all four equations.

Our problem, however, is not that we have too little information, but too much: By arbitrarily discarding one IV estimating equation from (4.13), we can obtain consistent estimators of the structural parameters. The surplus of instrumental variables indicates that the structural equation is overidentified.

The essential nature of overidentification is clarified by considering the population analogs of the estimating equations (4.13):

$$\begin{aligned}\sigma_{15} &= \gamma_{51}\sigma_{11} + \gamma_{52}\sigma_{12} + \beta_{56}\sigma_{16} \\ \sigma_{25} &= \gamma_{51}\sigma_{12} + \gamma_{52}\sigma_{22} + \beta_{56}\sigma_{26} \\ \sigma_{35} &= \gamma_{51}\sigma_{13} + \gamma_{52}\sigma_{23} + \beta_{56}\sigma_{36} \\ \sigma_{45} &= \gamma_{51}\sigma_{14} + \gamma_{52}\sigma_{24} + \beta_{56}\sigma_{46}\end{aligned}\tag{4.14}$$

If the model is correctly specified, and indeed the X 's and ϵ 's are independent, then equations (4.14) hold precisely and simultaneously.

It is helpful to think geometrically about the issue of overidentification. To simplify the geometry, imagine that we wish to estimate the structural equation

$$Y_5 = \gamma_{51}X_1 + \beta_{54}Y_4 + \epsilon_7\tag{4.15}$$

and have available three exogenous variables, X_1 , X_2 , and X_3 to serve as instruments. Applying these instrumental variables produces three estimating equations, each with a population analog. As we have pointed out with respect to the Duncan, Haller, and Portes model, if the model (4.15) is correctly specified, all three population equations hold simultaneously, despite the fact that there are but two unknown structural parameters. As depicted in Figure 4.8(a), each equation represents a line in the $\gamma_{51} \times \beta_{54}$ space. Since the equations hold simultaneously, the three lines intersect at a point, determining the true values of the parameters.

In the sample, however, the estimating equations are perturbed by sampling error; that is, while $\sigma_{\epsilon_7} = 0$, and while the average value (ignoring small-sample bias) of s_{ϵ_7} over many samples is zero, in a particular sample it is unlikely that s_{ϵ_7} is precisely zero. Geometrically, the lines corresponding to the three estimating equations do not in general intersect at a point, as shown in Figure 4.8(b), even if the model is correctly specified. Of course, if the estimating equations are highly inconsistent with one another (i.e., if the lines in Figure 4.8(b) enclose a large triangle), then we should suspect the specification of the model.

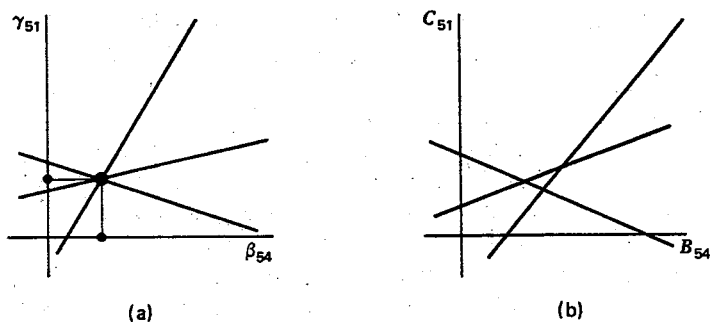


FIGURE 4.8. Overidentification in the population and the sample. (a) Population equations. (b) Sample estimating equations.

Returning to the peer-influences model, suppose that a path is added from X_3 to Y_5 , altering the first structural equation:

$$Y_5 = \gamma_{51}X_1 + \gamma_{52}X_2 + \gamma_{53}X_3 + \beta_{56}Y_6 + \varepsilon_7 \quad (4.16)$$

There are now four structural parameters to estimate. Because there are four instrumental variables available (the exogenous variables X_1 , X_2 , X_3 , and X_4), we may derive as many estimating equations as unknowns. We shall generally be able to solve uniquely for C_{51} , C_{52} , C_{53} , and B_{56} , and, therefore, the structural equation (4.16) is just identified. Note that we need not actually derive the IV estimating equations in order to draw this conclusion: We may simply compare the number of IVs to the number of parameters to be estimated, as in a balance sheet:

IVs ("credits")	Parameters ("debits")
X_1	γ_{51}
X_2	γ_{52}
X_3	γ_{53}
X_4	β_{56}
4	4

Now imagine that a path is added from X_4 to Y_5 (though this hardly makes substantive sense for the peer-influences data):

$$Y_5 = \gamma_{51}X_1 + \gamma_{52}X_2 + \gamma_{53}X_3 + \gamma_{54}X_4 + \beta_{56}Y_6 + \varepsilon_7 \quad (4.17)$$

Comparing the number of instrumental variables to the number of structural

parameters to be estimated indicates that there is a deficit of IVs:

IVs	Parameters
X_1	γ_{51}
X_2	γ_{52}
X_3	γ_{53}
X_4	γ_{54}
	β_{56}
4	5

There are five parameters to estimate, yet the pool of available instrumental variables yields but four estimating equations. The structural equation (4.17), therefore, is underidentified.

The *order condition* for identification is easily abstracted from these examples: For a structural equation to be identified, there must be at least as many exogenous variables (IVs) in the model as there are parameters to estimate in the structural equation.

The Admissible-Transformation Approach⁸ To understand why the order condition is insufficient to insure the identification of a structural equation, it is necessary to consider the equation not in isolation, but in relation to the other structural equations of the model. One way to accomplish this goal is to develop the relationship between the structural parameters and the parameters of the reduced form of the model.

In deriving the reduced form, we determined the following relation between reduced-form and structural parameters: $\Pi = -\mathbf{B}^{-1}\Gamma$ (see equation (4.4)). Knowing the structural parameters, then, we can find the reduced-form parameters. It is simple to show that, in general, it is not possible to reverse this relation: that is, to determine the structural parameters uniquely from the reduced form. Since the reduced form represents the directly observable empirical relationships of the endogenous to the exogenous variables, if two structures can give rise to the same reduced form, then it will be impossible to choose between these alternative structures on empirical grounds alone.

Let us multiply both sides of the structural equations (4.2) by a nonsingular ($q \times q$) transformation matrix \mathbf{T} , producing a new set of equations:

$$\mathbf{TB}y + \mathbf{T}\Gamma x = \mathbf{T}\epsilon \tag{4.18}$$

$$\mathbf{B}^*y + \Gamma^*x = \epsilon^*$$

where $\mathbf{B}^* \equiv \mathbf{TB}$, $\Gamma^* \equiv \mathbf{T}\Gamma$, and $\epsilon^* \equiv \mathbf{T}\epsilon$. Equation (4.18) only resembles a

⁸The general approach in this section is from Fisher (1966).

structural-equation model, because in general each row of \mathbf{B}^* and Γ^* combines parameters from *different* structural equations (different rows of \mathbf{B} and Γ). Yet, the reduced form corresponding to the "pseudo structure" (4.18) is the same as that corresponding to the "true structure" (4.2): Solving equation (4.18) for \mathbf{y} produces

$$\begin{aligned} \mathbf{y} &= -(\mathbf{TB})^{-1}\mathbf{TT}\mathbf{x} + (\mathbf{TB})^{-1}\mathbf{T}\boldsymbol{\varepsilon} \\ &= -\mathbf{B}^{-1}\mathbf{T}^{-1}\mathbf{TT}\mathbf{x} + \mathbf{B}^{-1}\mathbf{T}^{-1}\mathbf{T}\boldsymbol{\varepsilon} \\ &= -\mathbf{B}^{-1}\Gamma\mathbf{x} + \mathbf{B}^{-1}\boldsymbol{\varepsilon} \\ &= \Pi\mathbf{x} + \boldsymbol{\delta} \end{aligned}$$

The true and transformed structures, therefore, are *observationally indistinguishable* (i.e., they imply the same pattern of empirical relationships among the variables of the model).

We may, however, be able to distinguish true from transformed structures on the basis of the prior restrictions placed on the structural equations of the model. Suppose, for example, that some entries of Γ are prespecified to be zero. Then we may rule out any parameter matrix $\Gamma^* = \mathbf{TT}$ that contains nonzero entries where zeroes should appear. Following Fisher (1966), a transformation \mathbf{T} that produces a structure satisfying *all* prior restrictions placed on the model is termed an *admissible transformation*.

If we are able to show that the only admissible transformation is the identity transformation $\mathbf{T} = \mathbf{I}_q$, then the structural-equation model is identified. In fact, we need not be quite so stringent, for problems of underidentification only occur when we mix coefficients from different structural equations. It is therefore sufficient to require that the only admissible transformations are diagonal, multiplying each structural equation by a nonzero constant. Because the dependent variable in an equation has a coefficient of one, we can always recover the original structural equation by "renormalizing," dividing through by the same constant that we multiplied by (see footnote 4).

Before deriving a simple rule for determining whether the only admissible transformations of a structural-equation model are diagonal, let us consider two examples in some detail. First, we take another look at the Duncan, Haller, and Portes peer-influences model. There are no prior restrictions on \mathbf{B} in this model (other than the diagonal entries of one, which we have already taken account of), so any nonsingular transformation $\mathbf{B}^* = \mathbf{TB}$ is admissible from the point of view of \mathbf{B} . Γ , however, has four zero entries, and, therefore, if \mathbf{T} is an admissible transformation, then

$$\begin{aligned} \Gamma^* = \mathbf{TT} &= \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix} \begin{pmatrix} -\gamma_{51} & -\gamma_{52} & 0 & 0 \\ 0 & 0 & -\gamma_{63} & -\gamma_{64} \end{pmatrix} \\ &= \begin{pmatrix} -\gamma_{51}^* & -\gamma_{52}^* & 0 & 0 \\ 0 & 0 & -\gamma_{63}^* & -\gamma_{64}^* \end{pmatrix} \end{aligned} \quad (4.19)$$

Equation (4.19) requires that $t_{12} = t_{21} = 0$, and, thus, only diagonal transformations \mathbf{T} are admissible.

For a contrasting example, examine the model diagrammed in Figure 4.9. For this model,

$$\mathbf{B} = \begin{pmatrix} 1 & -\beta_{34} & 0 \\ -\beta_{43} & 1 & 0 \\ 0 & -\beta_{54} & 1 \end{pmatrix}$$

$$\mathbf{\Gamma} = \begin{pmatrix} -\gamma_{31} & 0 \\ 0 & 0 \\ 0 & -\gamma_{52} \end{pmatrix}$$

It is simple to show that there are nondiagonal admissible transformations that confuse the first structural equation with the second, despite the fact that all three structural equations of the model meet the order condition. That is,

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & 0 \\ 0 & t_{22} & 0 \\ 0 & 0 & t_{33} \end{pmatrix}$$

is admissible because

$$\mathbf{B}^* = \mathbf{TB} = \begin{pmatrix} t_{11} - t_{12}\beta_{43} & -t_{11}\beta_{34} + t_{12} & 0 \\ -t_{22}\beta_{43} & t_{22} & 0 \\ 0 & -t_{33}\beta_{54} & t_{33} \end{pmatrix}$$

$$\mathbf{\Gamma}^* = \mathbf{T}\mathbf{\Gamma} = \begin{pmatrix} -t_{11}\gamma_{31} & 0 \\ 0 & 0 \\ 0 & -t_{33}\gamma_{52} \end{pmatrix}$$

meet the prior restrictions placed on \mathbf{B} and $\mathbf{\Gamma}$. Upon renormalizing the first

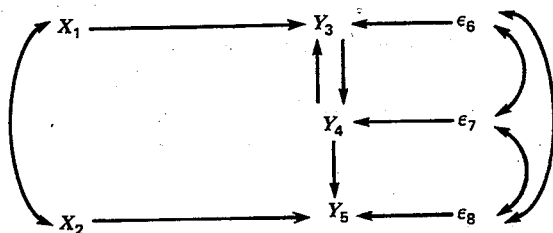


FIGURE 4.9. An underidentified nonrecursive model that meets the order condition.

equation, we obtain

$$1Y_3 - \frac{t_{11}\beta_{34} - t_{12}}{t_{11} - t_{12}\beta_{43}}Y_4 + 0Y_5 - \frac{t_{11}\gamma_{31}}{t_{11} - t_{12}\beta_{43}}X_1 + 0X_2 = \frac{t_{11}\epsilon_6 + t_{12}\epsilon_7}{t_{11} - t_{12}\beta_{43}}$$

$$1Y_3 - \beta_{34}^*Y_4 + 0Y_5 - \gamma_{31}^*X_1 + 0X_2 = \epsilon_6^*$$

Thus, if we obtain estimates for the first structural equation, we cannot be sure whether we have in fact estimated the structural parameters of interest or some confused combination of parameters from the first two structural equations. These difficulties occur because the second structural equation has zeroes in the same places as the first (the additional zero in equation two is irrelevant when we consider the identification of the first equation); if we take a linear combination of the first two equations, therefore, the zeroes still appear as specified in the prior restrictions on the first equation.

These observations suggest a relatively simple procedure for determining whether the only admissible transformations are diagonal. We examine each structural equation in turn, insuring that the corresponding row of T has zeroes except in the diagonal position. Without loss of generality, let us consider the first structural equation of a model. Let t'_1 denote the first row of T . The first structural equation is identified if and only if all entries except t_{11} in every admissible t'_1 are zero. Other entries in t'_1 may be nonzero only if linear combinations of the other equations meet the restrictions placed on the first structural equation.

We collect all structural coefficients of the model in a single matrix $A = [B, \Gamma]$. From this matrix, extract those columns that have zeroes in the first row, and then delete the first row, calling the resulting matrix A_1 . For example, for the Duncan, Haller, and Portes model, we have

$$A = \left(\begin{array}{cc|cc} 1 & -\beta_{56} & -\gamma_{51} & -\gamma_{52} & 0 & 0 \\ -\beta_{65} & 1 & 0 & 0 & -\gamma_{63} & -\gamma_{64} \end{array} \right)$$

$$A_1 = \begin{pmatrix} -\gamma_{63} & -\gamma_{64} \end{pmatrix}$$

More generally, A_1 is of order $(q - 1 \times r_1)$, where q is, as before, the number of equations in the model, and r_1 is the number of restrictions on (i.e., the number of variables excluded from) the first structural equation. If A_1 is of full-row rank (that is, if $\text{rank}(A_1) = q - 1$), then, by the definition of matrix rank, there is no linear combination of rows equalling the zero vector: The restrictions on the first structural equation cannot be duplicated from the other equations of the model. In the example, this is obviously the case since unless $\gamma_{63} = \gamma_{64} = 0$, $\text{rank}(A_1) = 1 = q - 1$.

Note that for the rank of A_1 to be $q - 1$, A_1 must have at least $q - 1$ columns. This is the order condition for identification in a new guise (and,

indeed, the name "order condition" refers to the column order of the A_1 matrix): There must be at least as many restrictions on a structural equation as one less than the number of endogenous variables in the model. That this condition is equivalent to the earlier statement of the order condition is easily verified by comparing the number of instrumental variables (i.e., exogenous variables) to the *potential* number of unknown parameters in a structural equation:

Number of IVs	Potential Number of Parameters
m	$m + q - 1$

There are $m + q - 1$ potential parameters because the coefficient of the dependent variable is fixed at one. Thus at least $q - 1$ potential independent variables must be excluded from a structural equation to reduce the number of parameters to or below the number of IVs:

$$r_1 \geq (m + q - 1) - m = q - 1$$

The *rank condition* must be met, of course, not just by the first structural equation, but by each structural equation of the model. That is,

$$\text{rank}(A_j) = q - 1$$

for $j = 1, \dots, q$. Only then are all admissible transformations diagonal. In practice, for models of the form considered here, the rank condition will be met if the order condition is met and if no structural equation duplicates the restrictions placed on any other.

4.3.2. The Identification of Recursive and Block-Recursive Models

As we shall show presently, the pool of instrumental variables for estimating an equation in a recursive or block-recursive model includes not only the exogenous variables but also prior endogenous variables. Put alternatively, the restrictions on disturbance covariances in recursive and block-recursive models may help to identify the models. We shall see, in fact, that all recursive models are identified.

Recall the recursive Blau and Duncan stratification model shown in Figure 4.2. The first structural equation of this model is

$$Y_3 = \gamma_{31} X_1 + \gamma_{32} X_2 + \varepsilon_6 \quad (4.20)$$

This equation has two parameters to be estimated, and two instrumental variables are available for estimation, the exogenous variables X_1 and X_2 . The structural equation (4.20), therefore, is just identified. More generally, the first

structural equation in a recursive model can contain only exogenous independent variables, for there are no endogenous variables causally prior to the first. Because the exogenous variables are also available as IVs, if all exogenous variables are included in the first equation, then the equation is just identified; if some exogenous variables are excluded, then the first structural equation is overidentified.

The second structural equation of the Blau and Duncan model is

$$Y_4 = \gamma_{42}X_2 + \beta_{43}Y_3 + \varepsilon_7 \quad (4.21)$$

Here, as in the first structural equation, there are two parameters to be estimated. The exogenous variables, as always, are available as IVs. In addition, the prior endogenous variable Y_3 may be used as an IV, since Y_3 is a linear combination of X_1 , X_2 , and ε_6 (as given in the first structural equation (4.20)), each of which is uncorrelated with the disturbance of the second structural equation, ε_7 : Because they are exogenous, X_1 and X_2 are uncorrelated with ε_7 ; ε_6 and ε_7 are uncorrelated because, in recursive models, different disturbance variables are specified to be independent. Y_3 , therefore, is uncorrelated with ε_7 . There are, then, three IVs for the structural equation (4.21), rendering this equation overidentified. Note that, more generally, the first endogenous variable in a recursive model is a linear combination of exogenous variables and a disturbance, and thus may be used as an IV in estimating the second equation of the model. This second equation, consequently, is overidentified if any prior variables are excluded and just identified otherwise.

Returning to our example, there are three parameters to be estimated in the third structural equation of the Blau and Duncan model:

$$Y_5 = \gamma_{52}X_2 + \beta_{53}Y_3 + \beta_{54}Y_4 + \varepsilon_8 \quad (4.22)$$

X_1 and X_2 may be employed as instrumental variables because they are exogenous. Y_3 is an eligible IV because it is composed of X_1 , X_2 , and ε_6 , each of which is uncorrelated with ε_8 . Y_4 , similarly, has components (X_2 , Y_3 , and ε_7) that are uncorrelated with ε_8 , and is a fourth IV. The structural equation (4.22) is consequently overidentified.

To generalize: Exogenous and all prior endogenous variables are eligible IVs for estimating a structural equation in a recursive model; if all of these variables are independent variables as well, the equation is just identified; if one or more prior variables are excluded, the equation is overidentified.⁹

Strictly speaking, we should set out to show that the only admissible transformations of a recursive model are diagonal. This may, in fact, be

⁹In certain cases, the causal ordering of endogenous variables in a recursive model is partial rather than complete. When this happens, each of two variables may count as "prior" relative to each other (i.e., uncorrelated with the disturbance in the other's equation). See Problem 4.8(a) for an example.

demonstrated by taking into account the restrictions placed on disturbance covariances as well as those placed on structural parameters (see Problem 4.9). Notice that in the context of recursive models, the rank condition derived in Section 4.3.1 is sufficient but not necessary to insure the identification of a model, for this condition fails to take account of restrictions on disturbance covariances.

An illustrative block-recursive model was given in Figure 4.4. There are four instrumental variables available for estimating the structural equations in the first block of this model: X_1 , X_2 , X_3 , and X_4 . Because each equation in the first block of the model has four parameters to be estimated, each equation is just identified. To estimate the structural equations in the second block, the pool of instrumental variables expands to include the endogenous variables Y_5 and Y_6 in the first block. This expansion occurs because Y_5 and Y_6 may be written (in reduced form) as linear functions of the exogenous variables and first-block errors, all of which are uncorrelated with the errors of the second block. Each structural equation in the second block has five parameters to be estimated, and therefore is overidentified. In the absence of the block-recursive restrictions on disturbance covariances, the second-block equations in this model would be *underidentified*.

In general, to identify a structural equation in a block-recursive model, we may employ endogenous variables in prior blocks (along with the exogenous variables) as IVs. A necessary and sufficient condition for identification may be obtained by suitably modifying the rank condition:

$$\text{rank}(\mathbf{A}_j^*) = q_j - 1$$

where \mathbf{A}_j^* is formed from \mathbf{A}_j by deleting equations (rows) for other blocks and by deleting endogenous variables (columns) in subsequent blocks. q_j is the number of endogenous variables in the block containing Y_j .

4.3.3. Restrictions on Disturbance Covariances: The General Case

The preceding section demonstrated how prior restrictions on disturbance covariances help to identify recursive and block-recursive structural-equation models. Restrictions of this type can also assist in identifying nonrecursive models, although there is not, unfortunately, a general rule (such as the order or rank condition) to apply in these cases. Instead, each model must be examined individually, employing, for example, the admissible-transformation approach.

We return to the general structural-equation model transformed by a nonsingular matrix \mathbf{T} (repeating equation (4.18)):

$$\mathbf{T}\mathbf{B}\mathbf{y} + \mathbf{T}\mathbf{T}\mathbf{x} = \mathbf{T}\boldsymbol{\varepsilon}$$

$$\mathbf{B}^*\mathbf{y} + \mathbf{T}^*\mathbf{x} = \boldsymbol{\varepsilon}^*$$

As we pointed out earlier, for \mathbf{T} to be an admissible transformation, \mathbf{B}^* and \mathbf{T}^*

must meet the prior restrictions placed on \mathbf{B} and Γ . Because ϵ^* results from a linear transformation of ϵ , the covariance matrix of the transformed disturbances may also be derived employing \mathbf{T} :

$$\Sigma_{\epsilon\epsilon}^* = \mathbf{T}\Sigma_{\epsilon\epsilon}\mathbf{T}'$$

For \mathbf{T} to be admissible, $\Sigma_{\epsilon\epsilon}^*$ must satisfy the prior restrictions placed on $\Sigma_{\epsilon\epsilon}$ (i.e., must have zeroes in the right places). We were able to ignore $\Sigma_{\epsilon\epsilon}^*$ previously because we stipulated that there were *no* restrictions placed on the disturbance covariance matrix.

To provide an illustration, we shall examine the model shown in Figure 4.10, the structural equations of which are

$$\begin{pmatrix} 1 & -\beta_{23} \\ -\beta_{32} & 1 \end{pmatrix} \begin{pmatrix} Y_2 \\ Y_3 \end{pmatrix} + \begin{pmatrix} 0 \\ -\gamma_{31} \end{pmatrix} X_1 = \begin{pmatrix} \epsilon_4 \\ \epsilon_5 \end{pmatrix}$$

(This model is adapted from Johnston, 1972: 365.) The two disturbance variables are specified to be uncorrelated:

$$\Sigma_{\epsilon\epsilon} = \begin{pmatrix} \sigma_{44} & 0 \\ 0 & \sigma_{55} \end{pmatrix}$$

Note that the second structural equation of the model would be underidentified if not for the restriction $\sigma_{45} = 0$.

To show that the model is in fact identified, it is necessary to demonstrate that only diagonal transformations are admissible. Let

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix}$$

There are no restrictions on \mathbf{B} , so any $\mathbf{B}^* = \mathbf{TB}$ will do. Because $\gamma_{21} = 0$, however, $t_{12} = 0$, for otherwise $\gamma_{21}^* = t_{11}0 - t_{12}\gamma_{31} \neq 0$, making \mathbf{T} inadmissible. Because $\sigma_{45} = 0$, the off-diagonal entries of $\Sigma_{\epsilon\epsilon}^*$ must be zero. Forming this product, we have

$$\begin{aligned} \Sigma_{\epsilon\epsilon}^* &= \mathbf{T}\Sigma_{\epsilon\epsilon}\mathbf{T}' = \begin{pmatrix} t_{11} & 0 \\ t_{21} & t_{22} \end{pmatrix} \begin{pmatrix} \sigma_{44} & 0 \\ 0 & \sigma_{55} \end{pmatrix} \begin{pmatrix} t_{11} & t_{21} \\ 0 & t_{22} \end{pmatrix} \\ &= \begin{pmatrix} t_{11}^2\sigma_{44} & t_{11}t_{21}\sigma_{44} \\ t_{11}t_{21}\sigma_{44} & t_{21}^2\sigma_{44} + t_{22}^2\sigma_{55} \end{pmatrix} \end{aligned}$$

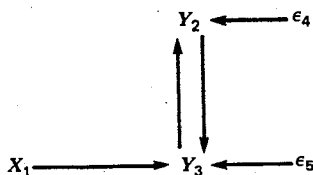


FIGURE 4.10. A nonrecursive model with restricted disturbance covariance.

It is clear that the off-diagonal entries are zero only if $t_{21} = 0$ (since $t_{11} \neq 0$ because \mathbf{T} is nonsingular). Thus, all admissible transformations are of the form

$$\mathbf{T} = \begin{pmatrix} t_{11} & 0 \\ 0 & t_{22} \end{pmatrix}$$

and the model is identified.

PROBLEMS

- 4.8. Determine the identification status of each of the models given in Problem 4.1.
- 4.9.* Using the admissible-transformation approach, demonstrate that a recursive model (i.e., a model for which \mathbf{B} is triangular and $\Sigma_{\epsilon\epsilon}$ is diagonal) is necessarily identified. (*Hints:* Employ the general approach used in Section 4.3.3. Work with a relatively simple system, such as Blau and Duncan's model modified so that all prior variables appear in each structural equation; the generalization to any recursive model is obvious.)

4.4. ESTIMATION OF STRUCTURAL-EQUATION MODELS

Having specified a structural-equation model, and having determined that it is identified, we wish to estimate the parameters of the model—that is, to fit the model to the data. If each of the structural equations of the model is just identified, estimation is not problematic, since the instrumental-variables estimating equations may then be solved for unique estimates of the structural parameters.

An overidentified model, however, presents difficulties: As we noted in Section 4.3.1, the sample estimating equations for an overidentified structural equation are overdetermined, even if the model is correctly specified. Recall the situation illustrated in Figure 4.8(b), where there are two parameters to be estimated and three estimating equations. To obtain consistent estimators of the structural parameters we could delete one of the estimating equations.

Though discarding surplus estimating equations is not an unreasonable response to overidentification, there are at least two factors that recommend against it. First, in the absence of a justifiable rule for determining which estimating equation is to be deleted, we must proceed arbitrarily, causing different investigators to obtain different estimates from the same data. Second, the surplus of information available when a model is overidentified might be used to improve the efficiency of estimation; discarding this information is statistically wasteful. In this section, we shall deal first with the estimation of nonrecursive models, and then turn to a consideration of recursive and block-recursive models.

4.4.1. Estimating Overidentified Nonrecursive Models

There are two general approaches to estimating an overidentified structural-equation model. One approach, called *single-equation* or *limited-information estimation*, is to estimate each structural equation separately. Ordinary least squares is an example of a single-equation method. OLS, however, is generally inconsistent when applied to nonrecursive models, because endogenous independent variables are correlated with the disturbance of the structural equation in which they appear. There are other single-equation methods that produce consistent estimators. One such method, *two-stage least squares (2SLS)*, is developed in this section.

A second general approach, termed *systems* or *full-information estimation*, estimates all of the parameters of the model (including the covariance matrix of the disturbances) at once. Later in this section, we shall take up the *full-information maximum-likelihood (FIML)* method.

Although a lengthy comparative discussion of different estimation methods is beyond the scope of this chapter, a few remarks are in order. Further summary information may be found in various sources, including Christ (1966: 464–481), Malinvaud (1970: 718–722), Kmenta (1971: 581–586), Johnston (1972: 408–420), and Wonnacott and Wonnacott (1979: 518–521).

The asymptotic (i.e., large-sample) properties of the various estimation methods have been determined analytically, and the full-information methods are asymptotically more efficient than the limited-information techniques. Determining the small-sample properties of structural-equation estimators analytically generally proves infeasible, and therefore these properties have primarily been explored empirically through “Monte-Carlo” (random-sampling) simulation experiments. The disadvantage of this approach, aside from its relative inelegance, is that conclusions may depend in an undetermined way on the specific conditions of a Monte-Carlo study (i.e., on the models, true parameter values, variable distributions, and sample sizes employed in the experiment). The random element in Monte-Carlo studies also introduces statistical uncertainty. Over the course of a number of studies, however, patterns tend to emerge.

To summarize the results of such studies briefly is possibly misleading. It is, nevertheless, fair to say that overall 2SLS appears to be the best of the limited-information methods, and FIML the best of the full-information methods. Moreover, FIML generally is superior to 2SLS, except when the estimated system has a misspecified equation, in which case single-equation methods like 2SLS tend to perform relatively well. Intuitively, full-information estimation proliferates a specification error throughout an equation system, while the limited-information approach isolates the error in a single equation.

A final point may be made with respect to OLS estimation. Though OLS is generally inconsistent in nonrecursive models, the OLS estimator nevertheless has smaller sampling variance than the consistent estimators. Even apart from the possible small-sample bias of the consistent estimators, their larger sam-

pling variance may depress their small-sample efficiency below that of OLS. OLS estimation, therefore, cannot simply be dismissed in very small samples on the grounds of inconsistency.¹⁰

Two-Stage Least Squares (2SLS) Aside from its desirable properties as an estimation method, 2SLS is worth studying because its rationale is relatively simple, it is computationally straightforward and inexpensive, and it is the method most frequently employed in practice. The 2SLS method was originally formulated in the 1950s by Theil (cited in Theil, 1971: 452) and Basman (1957). We shall approach 2SLS by developing an example before proceeding to the general case.

We return to the first structural equation of the Duncan, Haller, and Portes peer-influences model (originally given in equations (4.1)):

$$Y_5 = \gamma_{51}X_1 + \gamma_{52}X_2 + \beta_{56}Y_6 + \varepsilon_7 \quad (4.23)$$

Recall that this structural equation is overidentified because of the exclusion of the exogenous variables X_3 and X_4 . We therefore have four instrumental variables but only three structural parameters to estimate. 2SLS may be thought of as a method for reducing the number of instrumental variables to the number of parameters.

Aside from being uncorrelated with the error, a good instrumental variable should be as highly correlated as possible with the independent variables in the equation to be estimated. We may apply this criterion individually to the independent variables in equation (4.23). X_1 and X_2 , being perfectly correlated with themselves, are therefore their own best instruments. We might choose as an instrument for Y_6 the remaining exogenous variable (X_3 or X_4) that has the higher correlation with Y_6 . We can do better, however, by regressing Y_6 on *both* X_3 and X_4 , using the fitted values \hat{Y}_6 that result as an optimal instrument for Y_6 .

An equivalent, and ultimately more convenient, result is obtained by regressing Y_6 on *all* of the exogenous variables (X_1 , X_2 , X_3 , and X_4) in the model.¹¹ That is, in the first stage of 2SLS, we fit the reduced-form relation

$$Y_6 = \pi_{61}X_1 + \pi_{62}X_2 + \pi_{63}X_3 + \pi_{64}X_4 + \delta_6$$

¹⁰"Small samples" employed in econometric Monte-Carlo studies are small indeed by sociological standards—often in the neighborhood of 20. This is because economists frequently work with relatively short time series rather than with cross-sectional sample surveys. A social scientist employing a sample of several hundred or more observations almost certainly can rely on asymptotic results, if the assumptions underlying the results (e.g., the independence of exogenous variables and disturbances) are realistic.

¹¹Recall from Problem 4.7 that what is significant about a set of IVs is the subspace that they span, not the basis selected for this subspace. In the present context, since X_1 and X_2 are included among the set of IVs employed, it does not matter whether \hat{Y}_6 is defined in terms of all four X 's or only in terms of X_3 and X_4 ; the same subspace is spanned by X_1 , X_2 , and \hat{Y}_6 in both cases.

obtaining (from OLS estimates of the π 's)

$$\hat{Y}_6 = P_{61}X_1 + P_{62}X_2 + P_{63}X_3 + P_{64}X_4$$

\hat{Y}_6 , as a linear combination of the exogenous variables, is uncorrelated with the structural error ε_7 of equation (4.23), and, therefore, may legitimately be used as an instrumental variable in estimating this equation.¹² Moreover, \hat{Y}_6 is as highly correlated as possible with Y_6 while still remaining uncorrelated with ε_7 .

In the second stage of 2SLS, we apply the IVs obtained in the first stage (the exogenous independent variables X_1 and X_2 , and the fitted endogenous independent variable \hat{Y}_6) to estimate the structural equation (4.23). Using an obvious notation, we derive IV estimating equations

$$\begin{aligned} S_{15} &= C_{51}S_{11} + C_{52}S_{12} + B_{56}S_{16} \\ S_{25} &= C_{51}S_{12} + C_{52}S_{22} + B_{56}S_{26} \\ S_{56} &= C_{51}S_{16} + C_{52}S_{26} + B_{56}S_{66} \end{aligned} \quad (4.24)$$

which, given data, may be solved for 2SLS estimates of the structural parameters.

The name "two-stage least squares" derives from an alternative but equivalent approach employing an OLS regression in the second stage. The structural equation (4.23) could be fit directly by OLS but for the correlation of Y_6 with ε_7 . If we substitute for Y_6 from the first-stage reduced-form regression, we get

$$\begin{aligned} Y_5 &= \gamma_{51}X_1 + \gamma_{52}X_2 + \beta_{56}(\hat{Y}_6 + D_6) + \varepsilon_7 \\ &= \gamma_{51}X_1 + \gamma_{52}X_2 + \beta_{56}\hat{Y}_6 + \varepsilon_7^* \end{aligned} \quad (4.25)$$

Here, $\varepsilon_7^* \equiv \beta_{56}D_6 + \varepsilon_7$ is a linear combination of errors, and hence is uncorrelated with X_1 and X_2 , which are exogenous, and with \hat{Y}_6 , which is a linear combination of exogenous variables (see footnote 12). OLS estimation, therefore, may justifiably be applied to equation (4.25), producing estimating equations

$$\begin{aligned} S_{15} &= C_{51}S_{11} + C_{52}S_{12} + B_{56}S_{16} \\ S_{25} &= C_{51}S_{12} + C_{52}S_{22} + B_{56}S_{26} \\ S_{56} &= C_{51}S_{16} + C_{52}S_{26} + B_{56}S_{66} \end{aligned} \quad (4.26)$$

Comparing the OLS estimating equations (4.26) with the IV estimating equa-

¹² \hat{Y}_6 depends upon the P 's, which, in turn, depend upon the structural error ε_7 (a component of δ_6). \hat{Y}_6 and ε_7 , therefore, are not, strictly speaking, independent. Since the P 's are consistent estimators of the π 's, however, \hat{Y}_6 and ε_7 are independent in the limit.

tions (4.24) for the second stage, we need to show that $S_{16} = S_{1\bar{6}}$, $S_{26} = S_{2\bar{6}}$, and $S_{6\bar{6}} = S_{\bar{6}\bar{6}}$, in order to demonstrate the equivalence of the two approaches. This equivalence will be proven shortly, but for the general case, to which we now turn.

Let us consider the j th structural equation in a model, writing this equation in the following format:

$$y_j = Y_j \beta_j + X_j \gamma_j + \epsilon_j \quad (4.27)$$

$(n \times 1) \quad (n \times q_j - 1) \quad (q_j - 1 \times 1) \quad (n \times m_j) \quad (m_j \times 1) \quad (n \times 1)$

The symbols used in this equation, some of which are familiar, are explained in Table 4.1, which employs the first structural equation of the peer-influences model as an illustration. Equation (4.27) may be written more compactly as

$$y_j = [Y_j, X_j] \begin{pmatrix} \beta_j \\ \gamma_j \end{pmatrix} + \epsilon_j \quad (4.28)$$

In the first stage of 2SLS, we regress the endogenous independent variables in Y_j on *all* of the exogenous variables in the model:

$$Y_j = X \Pi_j' + \Delta_j$$

$(n \times q_j - 1) \quad (n \times m) \quad (m \times q_j - 1) \quad (n \times q_j - 1)$

TABLE 4.1. Notation for a Structural Equation to be Estimated by 2SLS

Symbol	Meaning	Example ^a
j	j th structural equation	1
n	Number of observations	329
q_j	Number of included endogenous variables	2
m_j	Number of exogenous independent variables	2
y_j	Dependent variable vector	y_5
$(n \times 1)$ Y_j	Endogenous independent variable matrix	y_6
$(n \times q_j - 1)$ X_j	Exogenous independent variable matrix	$[x_1, x_2]$
$(n \times m_j)$ ϵ_j	Disturbance vector	ϵ_7
$(n \times 1)$ β_j	Structural parameters for endogenous independent variables	(β_{56})
$(q_j - 1 \times 1)$ γ_j	Structural parameters for exogenous independent variables	$(\gamma_{51}, \gamma_{52})'$
$(m_j \times 1)$		

^aFirst equation of Duncan, Haller, and Portes nonrecursive model.

obtaining the OLS reduced-form estimator

$$P_j' = (X'X)^{-1}X'Y_j$$

and fitted values

$$\hat{Y}_j = XP_j' = X(X'X)^{-1}X'Y_j \quad (4.29)$$

In the second stage, we apply \hat{Y}_j and X_j as instrumental variables to equation (4.28):

$$\begin{aligned} \begin{pmatrix} b_j \\ c_j \end{pmatrix} &= ([\hat{Y}_j, X_j]'[Y_j, X_j])^{-1}[\hat{Y}_j, X_j]'y_j \\ &= \begin{pmatrix} \hat{Y}_j'Y_j & \hat{Y}_j'X_j \\ X_j'Y_j & X_j'X_j \end{pmatrix}^{-1} \begin{pmatrix} \hat{Y}_j'y_j \\ X_j'y_j \end{pmatrix} \end{aligned} \quad (4.30)$$

Alternatively, we may proceed by the regression approach, substituting $Y_j = \hat{Y}_j + D_j$ into equation (4.27):

$$\begin{aligned} y_j &= (\hat{Y}_j + D_j)\beta_j + X_j\gamma_j + \varepsilon_j \\ &= \hat{Y}_j\beta_j + X_j\gamma_j + (D_j\beta_j + \varepsilon_j) \\ &= [\hat{Y}_j, X_j] \begin{pmatrix} \beta_j \\ \gamma_j \end{pmatrix} + \varepsilon_j^* \end{aligned} \quad (4.31)$$

where $\varepsilon_j^* \equiv D_j\beta_j + \varepsilon_j$. We may apply OLS to equation (4.31) because, by the reasoning outlined earlier, \hat{Y}_j and X_j are both uncorrelated (in the limit) with ε_j^* :

$$\begin{aligned} \begin{pmatrix} b_j \\ c_j \end{pmatrix} &= ([\hat{Y}_j, X_j]'[\hat{Y}_j, X_j])^{-1}[\hat{Y}_j, X_j]'y_j \\ &= \begin{pmatrix} \hat{Y}_j'\hat{Y}_j & \hat{Y}_j'X_j \\ X_j'\hat{Y}_j & X_j'X_j \end{pmatrix}^{-1} \begin{pmatrix} \hat{Y}_j'y_j \\ X_j'y_j \end{pmatrix} \end{aligned} \quad (4.32)$$

It is clear that the two approaches, equations (4.30) and (4.32), produce identical results if

$$\begin{aligned} \hat{Y}_j'Y_j &= \hat{Y}_j'\hat{Y}_j \\ X_j'Y_j &= X_j'\hat{Y}_j \end{aligned} \quad (4.33)$$

To see that these equations hold, substitute $\mathbf{Y}_j = \hat{\mathbf{Y}}_j + \mathbf{D}_j$ into the left side of each, obtaining

$$\hat{\mathbf{Y}}_j'(\hat{\mathbf{Y}}_j + \mathbf{D}_j) = \hat{\mathbf{Y}}_j'\hat{\mathbf{Y}}_j + \hat{\mathbf{Y}}_j'\mathbf{D}_j$$

$$\mathbf{X}_j'(\hat{\mathbf{Y}}_j + \mathbf{D}_j) = \mathbf{X}_j'\hat{\mathbf{Y}}_j + \mathbf{X}_j'\mathbf{D}_j$$

$\hat{\mathbf{Y}}_j$ and \mathbf{D}_j are, respectively, the fitted dependent-variable matrix and the residual matrix from an OLS regression; they are, therefore, orthogonal, since each column of $\hat{\mathbf{Y}}_j$ lies in the \mathbf{X} subspace, and each column of \mathbf{D}_j is orthogonal to this subspace. For a similar reason, \mathbf{X}_j (a column subset of \mathbf{X}) and \mathbf{D}_j are orthogonal. Thus, $\hat{\mathbf{Y}}_j'\mathbf{D}_j$ and $\mathbf{X}_j'\mathbf{D}_j$ both vanish, and the equalities in (4.33) are demonstrated.

The estimated asymptotic covariance matrix of the 2SLS estimator follows from the observation that 2SLS is a type of instrumental-variables estimation. We may, consequently, apply the result given in Section 4.2, equations (4.11). In the present context, y_j plays the role of the dependent variable (y , in the general case given in (4.11)), $[\mathbf{Y}_j, \mathbf{X}_j]$ is the independent-variable matrix (\mathbf{X} in equations (4.11)), and $[\hat{\mathbf{Y}}_j, \mathbf{X}_j]$ is the matrix of instrumental variables (\mathbf{Z} in equations (4.11)). Straightforward substitution produces the desired result:

$$\mathbf{e}_j = \mathbf{y}_j - \mathbf{X}_j\mathbf{c}_j - \mathbf{Y}_j\mathbf{b}_j$$

$$S_{E_j}^2 = \frac{\mathbf{e}_j'\mathbf{e}_j}{n - q_j - m_j} \quad (4.34)$$

$$\widehat{\text{var}} \begin{pmatrix} \mathbf{b}_j \\ \mathbf{c}_j \end{pmatrix} = S_{E_j}^2 ([\hat{\mathbf{Y}}_j, \mathbf{X}_j]' [\mathbf{Y}_j, \mathbf{X}_j])^{-1} ([\hat{\mathbf{Y}}_j, \mathbf{X}_j]' [\hat{\mathbf{Y}}_j, \mathbf{X}_j]) ([\mathbf{Y}_j, \mathbf{X}_j]' [\mathbf{Y}_j, \mathbf{X}_j])^{-1}$$

This expression may be simplified by multiplying out the partitioned matrices and taking advantage of the identities given in equations (4.33):

$$\widehat{\text{var}} \begin{pmatrix} \mathbf{b}_j \\ \mathbf{c}_j \end{pmatrix} = S_{E_j}^2 \begin{pmatrix} \hat{\mathbf{Y}}_j'\mathbf{Y}_j & \mathbf{Y}_j'\mathbf{X}_j \\ \mathbf{X}_j'\mathbf{Y}_j & \mathbf{X}_j'\mathbf{X}_j \end{pmatrix}^{-1}$$

The square roots of the diagonal entries of this matrix are the standard errors of the estimated structural coefficients, which may be used, therefore, to test hypotheses and construct confidence intervals for individual β 's and γ 's.

Although we have developed 2SLS as a two-step procedure, the first stage (4.29) may be substituted into the second (4.30), bypassing separate calculation

of the first-stage regression:

$$\begin{pmatrix} \mathbf{b}_j \\ \mathbf{c}_j \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_j' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}_j & \mathbf{Y}_j' \mathbf{X}_j \\ \mathbf{X}_j' \mathbf{Y}_j & \mathbf{X}_j' \mathbf{X}_j \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}_j' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_j \\ \mathbf{X}_j' \mathbf{y}_j \end{pmatrix} \quad (4.35)$$

In fact, because of the location of matrix inverses in equation (4.35), every sum-of-squares-and-products matrix in this equation may be replaced by the corresponding sample covariance matrix:

$$\begin{pmatrix} \mathbf{b}_j \\ \mathbf{c}_j \end{pmatrix} = \begin{pmatrix} \mathbf{S}_{Y_j X} \mathbf{S}_{XX}^{-1} \mathbf{S}_{XY_j} & \mathbf{S}_{Y_j X_j} \\ \mathbf{S}_{X_j Y_j} & \mathbf{S}_{X_j X_j} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{S}_{Y_j X} \mathbf{S}_{XX}^{-1} \mathbf{s}_{X_j y_j} \\ \mathbf{s}_{X_j y_j} \end{pmatrix} \quad (4.36)$$

where, for example, $\mathbf{S}_{Y_j X} = [1/(n-1)] \mathbf{Y}_j' \mathbf{X}$. For models with standardized variables, the covariances in equation (4.36) are, of course, correlations.

A correlation matrix for the peer-influences data is given in Table 4.2. (This correlation matrix includes variables not employed in the current example, but which will be used later.) 2SLS estimates of the standardized structural

TABLE 4.2. Correlation Matrix for Peer-Influences Data, $n = 329$

	1	2	3	4	5	6	7	8	9
2	.6247								
3	.3269	.3669							
4	.4216	.3275	.6404						
5	.2137	.2742	.1124	.0839					
6	.4105	.4043	.2903	.2598	.1839				
7	.3240	.4047	.3054	.2786	.0489	.2220			
8	.2930	.2407	.4105	.3607	.0186	.1861	.2707		
9	.2995	.2863	.5191	.5007	.0782	.3355	.2302	.2950	
10	.0760	.0702	.2784	.1988	.1147	.1021	.0931	-.0438	.2087

- 1 Respondent's occupational aspiration score
- 2 Respondent's educational aspiration score
- 3 Best friend's occupational aspiration score
- 4 Best friend's educational aspiration score
- 5 Respondent's parental aspiration
- 6 Respondent's intelligence
- 7 Respondent's family socio-economic status (SES)
- 8 Best friend's family SES
- 9 Best friend's intelligence
- 10 Best friend's parental aspiration

Source: Duncan, Haller, and Portes (1968: Table 1). Reprinted from the *American Journal of Sociology* by permission of the University of Chicago Press. Copyright 1968, the University of Chicago Press.

parameters of the model are shown with parenthetical standard errors in equations (4.37):

$$Y_5 = 0.403 Y_6 + 0.272 X_1 + 0.151 X_2 + 0.841 E_7' \quad (4.37)$$

(0.104) (0.053) (0.054)

$$Y_6 = 0.341 Y_5 + 0.157 X_3 + 0.352 X_4 + 0.805 E_8'$$

(0.125) (0.054) (0.055)

The coefficients associated with the estimated disturbances require comment. It is usual practice, in a standardized model, to set the standard deviation of the estimated disturbance variables to one. Here, for example, $E_7' \equiv E_7/S_7$. By way of compensation, S_7 , often called a *residual path*, becomes the coefficient of E_7' in the estimated structural equation.

The estimates for the peer-influences model appear generally reasonable. All structural coefficients are positive, as expected, and corresponding coefficients in the two structural equations are similar in magnitude. It is, perhaps, surprising that the estimated structural coefficients for peer influences, B_{56} and B_{65} , are as large as they are relative to the coefficients for intelligence (C_{51} and C_{64}) and for SES (C_{52} and C_{63}).

Having obtained 2SLS estimates of the structural parameters of a model, we are generally interested in estimating as well the covariances among disturbances from different structural equations, if only to verify that these values are reasonable. Equation (4.34) gives us the variance of each estimated disturbance. Their covariances can be obtained similarly, say $S_{E_i E_j} = \mathbf{e}_i' \mathbf{e}_j / n$ (disregarding degrees of freedom). It is generally simpler, however, to compute the covariance matrix for the disturbances directly from structural coefficients and from covariances of endogenous and exogenous variables. According to the general structural-equation model (4.2), $\boldsymbol{\varepsilon} = \mathbf{B}\mathbf{y} + \Gamma\mathbf{x}$. Thus, since $E(\boldsymbol{\varepsilon}) = \mathbf{0}$,

$$\begin{aligned} \Sigma_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} &= E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = E[(\mathbf{B}\mathbf{y} + \Gamma\mathbf{x})(\mathbf{B}\mathbf{y} + \Gamma\mathbf{x})'] \\ &= \mathbf{B}E(\mathbf{y}\mathbf{y}')\mathbf{B}' + \mathbf{B}E(\mathbf{y}\mathbf{x}')\Gamma' + \Gamma E(\mathbf{x}\mathbf{y}')\mathbf{B}' + \Gamma E(\mathbf{x}\mathbf{x}')\Gamma' \\ &= \mathbf{B}\Sigma_{YY}\mathbf{B}' + \mathbf{B}\Sigma_{YX}\Gamma' + \Gamma\Sigma_{XY}\mathbf{B}' + \Gamma\Sigma_{XX}\Gamma' \end{aligned}$$

The result we are seeking follows upon substituting sample covariances and estimated structural coefficients for their population counterparts:

$$\mathbf{S}_{EE} = \mathbf{B}\mathbf{S}_{YY}\mathbf{B}' + \mathbf{B}\mathbf{S}_{YX}\mathbf{C}' + \mathbf{C}\mathbf{S}_{XY}\mathbf{B}' + \mathbf{C}\mathbf{S}_{XX}\mathbf{C}' \quad (4.38)$$

Estimated disturbance correlations may be calculated from the covariances in equation (4.38) in the usual manner; that is,

$$r_{E_i E_j} = \frac{S_{E_i E_j}}{S_{E_i} S_{E_j}}$$

For the Duncan, Haller, and Portes model estimated in equations (4.37), the disturbance correlation is $r_{78} = -.476$. A negative correlation between disturbances makes little substantive sense in this case, for we expect similar omitted causes of respondent's and best friend's aspirations. The negative correlation between disturbances, therefore, casts doubt upon the specification of the model. This point is developed in Gillespie and Fox (1980), and is pursued briefly in Section 4.6.2.

Full-Information Maximum Likelihood (FIML) It is surprising that the full-information maximum-likelihood method of estimation (Koopmans, Rubin, and Leipnik, 1950) antedates simpler estimation methods such as 2SLS. Application of FIML was not generally practical, however, until electronic computers became available to take on the formidable computational burden imposed by the method.

The derivation of the FIML estimator follows the usual maximum-likelihood approach.¹³ We begin with the general structural-equation model [from equation (4.2)]

$$B\mathbf{y}_i + \Gamma\mathbf{x}_i = \boldsymbol{\varepsilon}_i \quad (4.39)$$

and with the following distributional assumptions regarding the errors:

$$\boldsymbol{\varepsilon}_i \sim N_q(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}})$$

$$\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j \text{ independent for } i \neq j$$

$$\mathbf{x}_i, \boldsymbol{\varepsilon}_i \text{ independent}$$

From the formula for the multivariate-normal distribution, we have

$$p(\boldsymbol{\varepsilon}_i) = \frac{1}{(2\pi)^{q/2} |\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{\varepsilon}_i' \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^{-1} \boldsymbol{\varepsilon}_i\right) \quad (4.40)$$

We cannot apply equation (4.40) directly, because the disturbance vector $\boldsymbol{\varepsilon}_i$ is unobservable. We may, however, use the model (4.39) to transform $\boldsymbol{\varepsilon}_i$ to \mathbf{y}_i ,

¹³For an alternative approach to the derivation of the FIML estimator, see Christ (1966: 395-405). It is possible, moreover, to arrive at the same estimator by applying a heuristic variance-minimizing criterion (Wonnacott and Wonnacott, 1979: 521-526), providing a justification for the FIML estimator without the necessity for making the strong distributional assumptions required by the maximum-likelihood method.

treating \mathbf{x}_i as conditionally fixed and employing the Jacobian of the transformation:

$$\begin{aligned} p(\mathbf{y}_i|\mathbf{x}_i) &= p(\boldsymbol{\varepsilon}_i) \left| \frac{\partial \boldsymbol{\varepsilon}_i}{\partial \mathbf{y}_i} \right|_+ \\ &= p(\boldsymbol{\varepsilon}_i) \left| \frac{\partial(\mathbf{B}\mathbf{y}_i + \Gamma\mathbf{x}_i)}{\partial \mathbf{y}_i} \right|_+ \\ &= p(\boldsymbol{\varepsilon}_i) |\mathbf{B}|_+ \\ &= \frac{|\mathbf{B}|_+}{(2\pi)^{q/2} |\boldsymbol{\Sigma}_{\varepsilon\varepsilon}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{B}\mathbf{y}_i + \Gamma\mathbf{x}_i)' \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1} (\mathbf{B}\mathbf{y}_i + \Gamma\mathbf{x}_i)\right] \end{aligned}$$

Because the n observations are independent, their joint probability density conditional on the exogenous variables is given by the product of their marginal probability densities:

$$p(\mathbf{Y}|\mathbf{X}) = \frac{|\mathbf{B}|_+^n}{(2\pi)^{nq/2} |\boldsymbol{\Sigma}_{\varepsilon\varepsilon}|^{n/2}} \exp\left[-\frac{1}{2} \sum_{i=1}^n (\mathbf{B}\mathbf{y}_i + \Gamma\mathbf{x}_i)' \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1} (\mathbf{B}\mathbf{y}_i + \Gamma\mathbf{x}_i)\right]$$

and the logarithm of the likelihood function is

$$\begin{aligned} \log L(\mathbf{B}, \Gamma, \boldsymbol{\Sigma}_{\varepsilon\varepsilon}) &= n \log |\mathbf{B}|_+ - \frac{nq}{2} \log(2\pi) \\ &\quad - \frac{n}{2} \log |\boldsymbol{\Sigma}_{\varepsilon\varepsilon}| - \frac{1}{2} \sum_{i=1}^n (\mathbf{B}\mathbf{y}_i + \Gamma\mathbf{x}_i)' \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1} (\mathbf{B}\mathbf{y}_i + \Gamma\mathbf{x}_i) \end{aligned} \tag{4.41}$$

Note that the joint density for \mathbf{X} and \mathbf{Y} is given by $p(\mathbf{X}, \mathbf{Y}) = p(\mathbf{X})p(\mathbf{Y}|\mathbf{X})$. If the distribution of \mathbf{X} does not depend upon the parameters \mathbf{B} , Γ , and $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$, then maximizing $L(\mathbf{B}, \Gamma, \boldsymbol{\Sigma}_{\varepsilon\varepsilon})$ is equivalent to maximizing the joint likelihood for \mathbf{X} and \mathbf{Y} .

Because of the prior restrictions on the model, some of the entries of \mathbf{B} and Γ (and possibly of $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$) are constrained to be zero. Likewise, the diagonal entries of \mathbf{B} are fixed to one. Maximum-likelihood estimators of \mathbf{B} , Γ , and $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ maximize equation (4.41) subject to these constraints. The partial derivatives of the log likelihood with respect to the parameters are nonlinear, and therefore equation (4.41) must be maximized numerically. This is why FIML estimation is computationally burdensome.

As for maximum-likelihood estimation generally, the estimated asymptotic covariance matrix for the FIML estimator may be obtained from the inverse of

the information matrix evaluated at the estimated parameter values. Since maximum-likelihood estimators are asymptotically normally distributed, we may employ estimated standard errors for normal-distribution tests of the model parameters. Moreover, for an overidentified model, the general likelihood-ratio criterion yields a test of the overidentifying restrictions, as explained in Section 4.7.2.

FIML estimates for the Duncan, Haller, and Portes model are shown in equations (4.42):

$$\begin{aligned}
 Y_5 &= \underset{(0.053)}{0.237} X_1 + \underset{(0.047)}{0.176} X_2 + \underset{(0.104)}{0.398} Y_6 + 0.890E_7' \\
 Y_6 &= \underset{(0.047)}{0.219} X_3 + \underset{(0.056)}{0.311} X_4 + \underset{(0.131)}{0.422} Y_5 + 0.847E_8'
 \end{aligned}
 \tag{4.42}$$

These estimates are in reasonable agreement with the 2SLS estimates given in equations (4.37). The FIML method produces standard errors for estimated disturbance covariances, showing here that the embarrassing negative covariance between E_7 and E_8 is statistically highly significant: $S_{78} = -0.495$, with a standard error of 0.137.

4.4.2. Estimation of Recursive and Block-Recursive Models

In examining the identification status of recursive models (Section 4.3.2), we determined that all independent variables in a structural equation of a recursive model are uncorrelated with the disturbance of the equation. We may, therefore, consistently estimate any structural equation in a recursive model by OLS regression. Even if prior variables have been excluded from the equation in question, and there are consequently extra instrumental variables available, the Gauss-Markov theorem (Section 1.2.5) assures the optimality of the OLS estimator; by reasoning similar to that underlying 2SLS, each independent variable is its own best instrumental variable. Moreover, it may be shown that, for recursive models, the OLS and FIML estimators coincide (Land, 1973).

Correlations for the Blau and Duncan stratification data are shown in Table 4.3. OLS estimates for the Blau and Duncan recursive model appear in equations (4.43); standard errors are not given here because of the very large sample employed in Blau and Duncan's research.

$$\begin{aligned}
 Y_3 &= 0.310X_1 + 0.279X_2 + 0.859E_6' \\
 Y_4 &= 0.224X_2 + 0.440Y_3 + 0.818E_7' \\
 Y_5 &= 0.115X_2 + 0.394Y_3 + 0.281Y_4 + 0.753E_8'
 \end{aligned}
 \tag{4.43}$$

As for the nonrecursive peer-influences model, we have standardized the estimated disturbances, introducing residual paths into the structural equations. For a standardized structural equation estimated by OLS, the squared multiple correlation $R^2 = 1 - S_E^2$ (disregarding degrees of freedom for error). The model, therefore, accounts for 26.2, 33.1, and 43.3 percent of the variation in Y_3 , Y_4 , and Y_5 , consecutively. The estimated structural parameters are all positive, as expected, and assume reasonable values. In discussing the specification of the Blau and Duncan model, we were skeptical of the assumption that ε_7 and ε_8 are uncorrelated. A positive correlation between these disturbances would induce a positive correlation between Y_4 and ε_8 , tending to inflate B_{54} . This coefficient, however, is not strikingly large.

Block-recursive models may be estimated straightforwardly by IV, FIML, 2SLS, or some other applicable technique. Direct IV estimation may reasonably be undertaken for a just-identified structural equation; here, exogenous and prior endogenous variables comprise the pool of available instrumental variables. In applying FIML, we merely need to indicate the prior restrictions on $\Sigma_{\varepsilon\varepsilon}$ along with those on \mathbf{B} and Γ . For 2SLS estimation of an equation in a block-recursive model, endogenous variables in prior blocks should be treated as exogenous.

A block-recursive model for the peer-influences data was presented in Figure 4.4, and the correlations for the variables in this model were given in Table 4.2. 2SLS and FIML estimates for the model, along with their standard errors, appear in Table 4.4. Because the first two structural equations are just identified, the 2SLS estimates for these equations are the same as those obtained by direct application of the exogenous variables as instrumental variables. Furthermore, the 2SLS and FIML estimates for these equations are necessarily identical. Note that the two sets of estimates are similar, and that

TABLE 4.3. Correlations for Stratification Data, $n = 20,700$

	X_1	X_2	Y_3	Y_4	Y_5
X_1	1.000				
X_2	.516	1.000			
Y_3	.453	.438	1.000		
Y_4	.332	.417	.538	1.000	
Y_5	.322	.405	.596	.541	1.000

X_1 Father's education
 X_2 Father's occupational status
 Y_3 Education
 Y_4 First-job status
 Y_5 1962 occupational status

TABLE 4.4. Block-Recursive Model for the Duncan, Haller, and Portes Peer-Influences Data (Figure 4.4)

Coefficient for	Structural Equation for			
	Y_5	Y_6	Y_7	Y_8
<i>(a) FIML Estimates</i>				
X_1	0.2793 (0.0559) ^a	—	0.0939 (0.0397)	—
X_2	0.1535 (0.0559)	0.0772 (0.0599)	0.1865 (0.0462)	-0.0470 (0.0535)
X_3	0.0843 (0.0672)	0.2015 (0.0553)	-0.0398 (0.0491)	0.0697 (0.0480)
X_4	—	0.3574 (0.0567)	—	0.1589 (0.0436)
Y_5	—	0.2819 (0.1590)	0.4502 (0.0518)	—
Y_6	0.2804 (0.1362)	—	—	0.4202 (0.0522)
Y_7	—	—	—	0.3506 (0.0900)
Y_8	—	—	0.2235 (0.0875)	—
<i>(b) 2SLS Estimates</i>				
X_1	0.2793 (0.0563)	—	0.1391 (0.0475)	—
X_2	0.1535 (0.0562)	0.0772 (0.0603)	0.1864 (0.0470)	-0.0428 (0.0544)
X_3	0.0843 (0.0676)	0.2105 (0.0556)	-0.0367 (0.0499)	0.0707 (0.0489)
X_4	—	0.3574 (0.0571)	—	0.1825 (0.0522)
Y_5	—	0.2819 (0.1600)	0.4034 (0.0554)	—
Y_6	0.2804 (0.1371)	—	—	0.4063 (0.0576)
Y_7	—	—	—	0.3367 (0.0910)
Y_8	—	—	0.2078 (0.0876)	—

^aStandard errors in parentheses.

the unreasonable negative correlation between estimated disturbances appears in this model as well.

PROBLEMS

- 4.10. Indicate how you would estimate each structural-equation model specified in Problem 4.1.
- 4.11. Data for Rindfuss, Bumpass, and St. John's fertility model (Problem 4.2) are given in Table 4.5. Fit the model to these data using an appropriate estimation method. Comment on the results.
- 4.12. Table 4.6 shows a covariance matrix among the variables in Berk and Berk's model for the division of household labor (Problem 4.3). Use these covariances to estimate the model.
- 4.13. The correlations in Table 4.7 are for non-black men in the experienced civilian labor force, who were of non-farm background and 35-44 years old in 1962. Use these correlations to estimate Duncan, Featherman, and Duncan's stratification model (Problem 4.4).
- 4.14. The covariances in Table 4.8 were calculated from data presented by Lincoln. Using these covariances, estimate Lincoln's model for strike activity in metropolitan areas (Problem 4.5).

TABLE 4.5. Covariances for Rindfuss, Bumpass, and St. John's Fertility Data

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	456.6769							
X_2	-0.9201	0.0894						
X_3	-15.8253	0.1416	9.2112					
X_4	-3.2442	0.0124	0.3908	0.2209				
X_5	-1.3205	0.0451	0.2181	0.0491	0.2294			
X_6	-0.4631	0.0174	-0.0458	-0.0055	0.0132	0.1498		
X_7	0.4768	-0.0191	0.0179	-0.0295	-0.0589	0.0085	0.1772	
X_8	-0.3143	0.0031	0.0291	-0.0096	-0.0018	0.0089	-0.0014	0.1170
X_9	0.2356	0.0031	0.0018	-0.0045	-0.0039	0.0021	-0.0003	0.0009
X_{10}	18.6603	-0.1567	-2.3493	-0.2052	-0.2385	-0.1434	-0.0119	-0.1380
Y_{11}	16.2133	-0.2305	-1.4237	-0.2262	-0.3458	0.1752	0.1683	-0.1702
Mean	30.209	0.099	3.889	0.330	0.357	0.183	0.231	0.136
	X_9	Y_{10}	Y_{11}					
X_9	0.0888							
Y_{10}	0.0267	5.5696						
Y_{11}	0.2626	3.6580	16.6382					
Mean	0.099	11.595	22.012					

Source: Adapted with permission from Rindfuss, Bumpass, and St. John (1980: 436, 445).

TABLE 4.6. Covariances for Berk and Berk's Data on the Division of Household Labor

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	0.01440							
X_2	-0.00288	0.02250						
X_3	-0.00898	0.00382	0.02890					
X_4	-0.00664	0.01226	0.00538	0.06946				
X_5	-0.01673	0.02214	0.00348	0.05079				
X_6	-0.00433	0.00085	0.00807	0.00300	0.16810			
X_7	0.00281	-0.00135	-0.00520	0.00019	-0.00779	0.03610		
X_8	0.00228	0.00028	-0.00484	-0.00501	-0.00148	-0.00137	0.03240	
X_9	-0.00114	0.00456	-0.00420	0.00015	-0.00779	-0.00144	-0.00137	0.03610
X_{10}	0.00158	0.00033	-0.00636	-0.00638	0.00234	-0.00167	-0.00158	0.00334
X_{11}	0.00540	0.00225	-0.01122	-0.00949	-0.00984	-0.00342	-0.00324	0.01254
X_{12}	0.00437	-0.00195	-0.00707	-0.00274	-0.00426	-0.00247	-0.00234	0.00296
X_{13}	0.06754	0.19698	-0.52622	-0.09889	0.23075	0.03564	-0.33768	0.32080
X_{14}	-0.04208	-0.00928	0.01403	0.00815	0.04229	0.00784	-0.00743	-0.01960
X_{15}	-0.00360	0.01200	-0.00680	0.02504	0.06355	-0.00570	0.00180	-0.00066
X_{16}	-0.00027	-0.00270	0.00688	-0.01898	-0.03321	0.01026	0.00162	-0.00513
X_{17}	-0.00730	0.01104	0.00653	0.01434	0.01443	0.00669	-0.00346	-0.00426
X_{18}	-0.01214	-0.00828	0.03284	0.00242	-0.00189	0.01049	-0.00994	-0.01136
X_{19}	-0.00836	-0.00062	0.02927	0.00756	0.00336	0.00234	-0.00664	-0.00234
X_{20}	-0.00936	-0.00526	0.01392	0.00411	0.00799	0.00370	-0.00632	-0.00741
X_{21}	-0.01426	-0.00648	0.02570	0.00474	-0.00074	0.00479	-0.00518	-0.00616
X_{22}	-0.02534	0.05472	-0.01958	0.06072	0.17318	0.02189	0.00207	0.00730
X_{23}	-0.01732	0.00936	0.02254	0.00925	0.00320	0.00889	-0.00632	-0.00741
Mean	0.84	0.21	0.16	0.49	0.21	0.04	0.03	0.04
	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
X_9	0.03610							
X_{10}	-0.00167	0.04840						
X_{11}	0.00684	0.00066	0.09000					
X_{12}	-0.00247	0.00172	-0.00702	0.06760				
X_{13}	0.44555	0.39208	0.78792	0.48776	87.98440			
X_{14}	-0.02352	-0.02042	-0.00309	-0.01609	-0.29023	1.06372		
X_{15}	-0.00095	0.01100	0.00600	0.02340	1.31320	0.10829	0.25000	
X_{16}	-0.00513	-0.00891	0.00945	-0.00936	-0.21105	0.06962	-0.12825	0.20250
X_{17}	-0.00426	0.00563	-0.01152	0.00250	0.15008	-0.00990	0.00960	-0.00576
X_{18}	-0.00524	-0.00911	-0.02346	-0.01555	-0.86296	0.00474	-0.02760	0.01863
X_{19}	-0.00779	-0.01082	-0.02091	-0.01492	-0.07692	0.02960	-0.02460	0.02583
X_{20}	-0.00741	-0.00858	-0.01872	-0.01318	-1.06088	0.01609	-0.03510	-0.00035
X_{21}	-0.00616	-0.00792	-0.01512	-0.00562	-0.97927	0.04827	-0.02700	0.02430
X_{22}	0.01094	0.02112	0.08064	0.02995	3.60192	0.79209	0.53760	-0.01728
X_{23}	-0.00741	-0.00944	0.01872	-0.00811	-0.43898	0.06436	-0.00390	-0.00035
Mean	0.03	0.05	0.10	0.07	59.28	1.75	0.44	0.29
	X_{17}	X_{18}	X_{19}	X_{20}	Y_{21}	Y_{22}	Y_{23}	
X_{17}	0.10240							
X_{18}	0.01619	0.21160						
X_{19}	0.00262	0.01320	0.16810					
X_{20}	-0.00499	0.05203	-0.00160	0.15210				
Y_{21}	0.00346	0.01325	0.02066	0.00842	0.12960			
Y_{22}	0.00614	-0.04416	-0.11808	-0.06739	-0.04147	3.68640		
Y_{23}	0.05990	0.03050	0.01919	0.02282	0.00842	0.00749	0.15210	
Mean	0.11	0.29	0.27	0.19	0.16	4.94	0.19	

Source: Reprinted from Richard A. Berk and Sarah Fenstermaker Berk, "A Simultaneous Equation Model for the Division of Household Labor," *Sociological Methods and Research*, Vol. 6, No. 4 (May 1978), pp. 431-468, with permission of Sage Publications, Inc.

TABLE 4.7. Duncan, Featherman, and Duncan's Stratification Data

	X_1	X_2	X_3	Y_4	Y_5	Y_6
X_1	1.0000					
X_2	0.5300	1.0000				
X_3	-0.2871	-0.2476	1.0000			
Y_4	0.4048	0.4341	-0.3311	1.0000		
Y_5	0.3194	0.3899	-0.2751	0.6426	1.0000	
Y_6	0.2332	0.2587	-0.1752	0.3759	0.4418	1.0000

Source: Reprinted with permission from Duncan, Featherman, and Duncan (1972: Table 3.1).

TABLE 4.8. Covariances for Lincoln's Strike-Activity Data

	X_1	X_2	X_3	X_4	Y_5	Y_6	Y_7
X_1	0.007744						
X_2	0.000635	0.000400					
X_3	0.052401	0.005077	1.065024				
X_4	0.006624	0.001471	0.066069	0.037636			
Y_5	0.054564	0.012024	0.823108	0.137249	1.809025		
Y_6	0.084675	0.015990	1.131609	0.171958	2.025220	2.496400	
Y_7	0.103616	0.019572	1.325756	0.184820	1.969703	2.567911	2.989441
Mean	0.509	0.752	12.332	0.649	5.009	6.182	8.832

Source: Adapted with permission from Lincoln (1978: 208).

**4.5. PATH ANALYSIS OF RECURSIVE MODELS:
CAUSAL INFERENCE AND DATA ANALYSIS**

Path analysis is often taken to be synonymous with structural-equation modeling. We shall use the term in a more literal and delimited sense to mean the decomposition of statistical relationships between pairs of variables into causal and noncausal components.

In data analysis generally, when we assess the causal impact of one variable on another, we are motivated to control for some third variable or set of variables (in the absence of interaction effects) in two substantively different contexts: (1) the third variable (say, Z) intervenes causally between the other two (say, X and Y); (2) the third variable is a common prior cause of the other two. These two contrasting situations are illustrated in Figure 4.11. The broken arrows in this figure indicate that, in either causal structure, X may or may not exert a direct impact on Y .

Let us suppose, for the moment, that the direct effect is absent or negligible. It is perhaps surprising, but nonetheless true, that our statistical expectations in the two situations are identical: In both cases, the partial relationship between X and Y vanishes when we control for Z . Although the observable consequences of the two causal schemes are, then, identical, their interpretation

is importantly different: In the first case we have *explained* the mechanism according to which X affects Y (that is, through the intervening variable Z), while in the second case we have *explained away* the empirical association between X and Y as "spurious" (that is, due to the common prior cause Z , not to the effect of X on Y).

It is instructive to examine the causal schemes in Figure 4.11 in an explicit structural-equation setting. In each case, suppose that the structure is recursive; then the equation for Y is

$$Y = \beta_1 X + \beta_2 Z + \epsilon \quad (4.44)$$

Here we use β 's for structural coefficients despite the fact that X is exogenous in the first model and Z is exogenous in the second. Now imagine that instead of fitting equation (4.44), we fit

$$Y = \beta_1 X + \epsilon' \quad (4.45)$$

where $\epsilon' \equiv \epsilon + \beta_2 Z$. Including Z in the error will cause our OLS estimator of β_1 to be biased, since X and Z are correlated. That is,

$$\text{plim } B_1 = \frac{\sigma_{XY}}{\sigma_{XX}} = \beta_1 + \beta_2 \frac{\sigma_{XZ}}{\sigma_{XX}} \quad (4.46)$$

If, however, Z intervenes causally between X and Y [Figure 4.11(a)], the "bias" in equation (4.46) is simply the indirect effect of X on Y through Z . (Note that σ_{XZ}/σ_{XX} is the coefficient for the simple regression of Z on X ; since X is causally prior to Z , this coefficient represents the effect of X on Z .) In other words, fitting equation (4.45) in place of equation (4.44), failing to control for an intervening variable, produces correct, albeit simplified, conclusions. In contrast, if Z is causally prior to both X and Y , then the bias term in equation (4.46) represents a spurious source of association between X and Y . To fail to control for a common antecedent cause, therefore, is to commit an error of causal inference.

Two important related conclusions may be drawn from this discussion. First, we cannot expect our data to mediate issues of causal priority, since very different causal structures have the same observable implications. Second, the conclusions that we draw from a data analysis depend centrally on the causal relations that are assumed to hold among the "independent" variables in the analysis.

In introducing structural-equation models to sociologists, Duncan (1966) argued that one of the strengths of the method "is that it produces a calculus



FIGURE 4.11. Two causal structures. (a) Intervening variable. (b) Common prior cause.

for indirect effects." This topic has been taken up by a number of researchers, including Finney (1972), Lewis-Beck (1974), Alwin and Hauser (1975), Lewis-Beck and Mohr (1976), Greene (1977), Fox (1980), and Sobel (1982). The approach adopted here is from Fox (1980). There has been some discussion of path-analytic decompositions for nonrecursive models (Lewis-Beck and Mohr, 1976; Heise, 1975; Fox, 1980), but we shall limit consideration to recursive models. Although decompositional methods have typically been applied to standardized models, we shall develop the more general unstandardized case.

We begin by noting that a structural-equation model implies a set of covariances among the endogenous variables of the model, and between endogenous and exogenous variables. These covariances may be derived by the now familiar expectation method. Multiplying the structural equations (4.2) through on the right by x' and taking expectations, we get

$$\begin{aligned} BE(yx') + \Gamma E(xx') &= E(\epsilon x') \\ B\Sigma_{YX} + \Gamma\Sigma_{XX} &= 0 \\ \Sigma_{YX} &= -B^{-1}\Gamma\Sigma_{XX} \end{aligned} \tag{4.47}$$

Similarly,

$$\begin{aligned} \Sigma_{YY} &= E(yy') \\ &= E[(-B^{-1}\Gamma x + B^{-1}\epsilon)(-B^{-1}\Gamma x + B^{-1}\epsilon)'] \\ &= B^{-1}\Gamma\Sigma_{XX}\Gamma'(B^{-1})' + B^{-1}\Sigma_{\epsilon\epsilon}(B^{-1})' \end{aligned} \tag{4.48}$$

The entries of Σ_{YX} and Σ_{YY} are shown in scalar form for the Blau and Duncan stratification model (Figure 4.2) in equations (4.49):

$$\begin{aligned} \sigma_{13} &= \gamma_{31}\sigma_{11} + \gamma_{32}\sigma_{12} \\ \sigma_{23} &= \gamma_{31}\sigma_{12} + \gamma_{32}\sigma_{22} \\ \sigma_{14} &= \gamma_{42}\sigma_{12} + \beta_{43}\gamma_{31}\sigma_{11} + \beta_{43}\gamma_{32}\sigma_{12} \\ \sigma_{24} &= \gamma_{42}\sigma_{22} + \beta_{43}\gamma_{31}\sigma_{12} + \beta_{43}\gamma_{32}\sigma_{22} \\ \sigma_{34} &= \gamma_{42}\gamma_{31}\sigma_{12} + \gamma_{42}\gamma_{32}\sigma_{22} + \beta_{43}\sigma_{33} \\ \sigma_{15} &= \gamma_{52}\sigma_{12} + \beta_{53}\gamma_{31}\sigma_{11} + \beta_{53}\gamma_{32}\sigma_{12} \\ &\quad + \beta_{54}\gamma_{42}\sigma_{12} + \beta_{54}\beta_{43}\gamma_{31}\sigma_{11} + \beta_{54}\beta_{43}\gamma_{32}\sigma_{12} \\ \sigma_{25} &= \gamma_{52}\sigma_{22} + \beta_{53}\gamma_{31}\sigma_{12} + \beta_{53}\gamma_{32}\sigma_{22} \\ &\quad + \beta_{54}\gamma_{42}\sigma_{22} + \beta_{54}\beta_{43}\gamma_{31}\sigma_{12} + \beta_{54}\beta_{43}\gamma_{32}\sigma_{22} \\ \sigma_{35} &= \gamma_{52}\gamma_{31}\sigma_{12} + \gamma_{52}\gamma_{32}\sigma_{22} + \beta_{53}\sigma_{33} \\ &\quad + \beta_{54}\gamma_{42}\gamma_{31}\sigma_{12} + \beta_{54}\gamma_{42}\gamma_{32}\sigma_{22} + \beta_{54}\beta_{43}\sigma_{33} \\ \sigma_{45} &= \gamma_{52}\gamma_{42}\sigma_{22} + \gamma_{52}\beta_{43}\gamma_{31}\sigma_{12} + \gamma_{52}\beta_{43}\gamma_{32}\sigma_{22} \\ &\quad + \beta_{53}\gamma_{42}\sigma_{12} + \beta_{53}\gamma_{42}\gamma_{32}\sigma_{22} + \beta_{53}\beta_{43}\sigma_{33} + \beta_{54}\sigma_{44} \end{aligned} \tag{4.49}$$

Note that we have not written down expressions for the variances (σ_{33} , σ_{44} , and σ_{55}) of the endogenous variables, nor have we substituted for these variances when they appear on the right-hand side of the equations. The equations in (4.49) are obtained by multiplying each structural equation of the model by all prior variables, taking expectations, and successively substituting for all covariances involving an endogenous variable, until only structural parameters, exogenous covariances, and variances appear on the right-hand side. To decompose σ_{15} , for instance, we proceed as follows:

$$E(X_1 Y_5) = \gamma_{52} E(X_1 X_2) + \beta_{53} E(X_1 Y_3) + \beta_{54} E(X_1 Y_4)$$

$$\begin{aligned} \sigma_{15} &= \gamma_{52}\sigma_{12} + \beta_{53}\sigma_{13} + \beta_{54}\sigma_{14} \\ &= \gamma_{52}\sigma_{12} + \beta_{53}(\gamma_{31}\sigma_{11} + \gamma_{32}\sigma_{12}) + \beta_{54}(\gamma_{42}\sigma_{12} + \beta_{43}\gamma_{31}\sigma_{11} + \beta_{43}\gamma_{32}\sigma_{12}) \\ &= \gamma_{52}\sigma_{12} + \beta_{53}\gamma_{31}\sigma_{11} + \beta_{53}\gamma_{32}\sigma_{12} + \beta_{54}\gamma_{42}\sigma_{12} + \beta_{54}\beta_{43}\gamma_{31}\sigma_{11} + \beta_{54}\beta_{43}\gamma_{32}\sigma_{12} \end{aligned}$$

Dividing each covariance in equations (4.49) by the variance of the causally prior variable produces the population slope for the simple linear regression of each endogenous variable on each prior variable. For example, dividing σ_{25} by σ_{22} , and σ_{35} by σ_{33} , we obtain

$$\begin{aligned} \mu_{52} = \frac{\sigma_{25}}{\sigma_{22}} &= \gamma_{52} + \beta_{53}\gamma_{31} \frac{\sigma_{12}}{\sigma_{22}} + \beta_{53}\gamma_{32} \\ &\quad + \beta_{54}\gamma_{42} + \beta_{54}\beta_{43}\gamma_{31} \frac{\sigma_{12}}{\sigma_{22}} + \beta_{54}\beta_{43}\gamma_{32} \\ \mu_{53} = \frac{\sigma_{35}}{\sigma_{33}} &= \gamma_{52}\gamma_{31} \frac{\sigma_{12}}{\sigma_{33}} + \gamma_{52}\gamma_{32} \frac{\sigma_{22}}{\sigma_{33}} + \beta_{53} \\ &\quad + \beta_{54}\gamma_{42}\gamma_{31} \frac{\sigma_{12}}{\sigma_{33}} + \beta_{54}\gamma_{42}\gamma_{32} \frac{\sigma_{22}}{\sigma_{33}} + \beta_{54}\beta_{43} \end{aligned} \tag{4.50}$$

These simple-regression or "gross" slopes measure (in Goldberger's, 1973, terminology) the *empirical association* between each pair of variables.

Equations (4.50) illustrate how a gross slope may be decomposed into path components of three general types: (1) a *direct effect*, represented by a structural coefficient; (2) *indirect effects*, given by products of structural coefficients along a path linking a prior variable to an endogenous variable; and (3) *noncausal components*. For the association between an exogenous and endogenous variable, the noncausal components are termed *unanalyzed*, because they depend upon covariances among exogenous variables, for which a causal ordering is not distinguished. The noncausal components of the association between two endogenous variables are termed *spurious*, because they depend upon correlated causes of, or causes common to, both variables. Examples of the various sorts of components are given in Figure 4.12.

In the sample, we may estimate model-implied covariances by substituting for the population quantities appearing in equations (4.47) and (4.48):

$$S_{YX}^* = -B^{-1}CS_{XX} \tag{4.51}$$

$$S_{YY}^* = B^{-1}CS_{XX}C'(B^{-1})' + B^{-1}S_{EE}(B^{-1})'$$

S_{YX}^* and S_{YY}^* are starred because, in an overidentified model, they may differ from covariances calculated directly from the data (i.e., $S_{YX} = [1/(n-1)]Y'X$ and $S_{YY} = [1/(n-1)]Y'Y$). In a recursive model, B and C are obtained by OLS regression, and S_{EE} is diagonal. Sample model-implied gross slopes are then given by

$$M_{YX}^* = S_{YX}^*V_X^{-1}$$

$$M_{YY}^* = S_{YY}^*V_Y^{*-1}$$

where $V_X \equiv \text{diag}(S_{XX})$ and $V_Y^* \equiv \text{diag}(S_{YY}^*)$.

Let us denote by E_{YX} and E_{YY} , respectively, the matrices of total effects (direct and indirect) of the exogenous on the endogenous variables, and of the

	Prior Variable	
	Exogenous	Endogenous
Gross slope	$\mu_{52} = \sigma_{25}/\sigma_{22}$	$\mu_{53} = \sigma_{35}/\sigma_{33}$
Type of Component		
(1) Direct effect	γ_{52} $X_2 \rightarrow Y_5$	β_{53} $Y_3 \rightarrow Y_5$
(2) Indirect effect	$\beta_{54}\gamma_{42}$ $X_2 \rightarrow Y_4 \rightarrow Y_5$	$\beta_{54}\beta_{43}$ $Y_3 \rightarrow Y_4 \rightarrow Y_5$
(3) Noncausal		
(a) Unanalyzed	$\beta_{53}\gamma_{31}\sigma_{12}/\sigma_{22}$ $X_1 \rightarrow Y_3 \rightarrow Y_5$ X_2	—
(b) Spurious	—	$\gamma_{52}\gamma_{32}\sigma_{22}/\sigma_{33}$ $X_2 \rightarrow Y_3 \rightarrow Y_5$
(i) Common prior cause		
(ii) Correlated prior causes		$X_1 \rightarrow Y_3$ $X_2 \rightarrow Y_5$ $\beta_{53}\gamma_{31}\sigma_{12}/\sigma_{33}$

FIGURE 4.12. Examples of path components from the Blau and Duncan model. (Adapted from Fox, 1980: Figure 2.)

endogenous variables on each other. Fox (1980) shows that¹⁴

$$E_{YX} = -B^{-1}C$$

$$E_{YY} = B^{-1} - I_q$$

Notice that E_{YX} is simply the reduced-form coefficient matrix obtained from the estimated structural parameters. Noncausal components may be calculated by subtraction:

$$N_{YX} = M_{YX}^* - E_{YX}$$

$$N_{YY} = M_{YY}^* - E_{YY}$$

where N_{YX} and N_{YY} are matrices of unanalyzed and spurious components, respectively. In a recursive model, E_{YY} is lower triangular (a later variable cannot affect a causally prior one); thus, the upper triangle of M_{YY}^* is necessarily wholly spurious and normally would not be shown.

Direct effects are given by the structural coefficients themselves:

$$D_{YX} = -C$$

$$D_{YY} = -(B - I_q) = I_q - B$$

Indirect effects follow by subtraction:

$$I_{YX} = E_{YX} - D_{YX}$$

$$I_{YY} = E_{YY} - D_{YY}$$

An effect analysis for the standardized Blau and Duncan model appears in Table 4.9. (In applying the results of this section to a standardized model, it is merely necessary to substitute correlations for covariances.) Some of the information in Table 4.9 has been reorganized in Table 4.10 to show the sources of association between each prior variable in the Blau and Duncan model and the final endogenous variable, 1962 occupation, Y_5 . The last column in the table, labeled B^* , gives the standardized partial regression coefficients for the regression of Y_5 on all prior variables; although this regression does not follow from the Blau and Duncan model, which sets $\gamma_{51} = 0$ *a priori*, we shall shortly have occasion to make reference to these coefficients. Note, for the present, that the regression coefficients in the last column of the table are

¹⁴Fox treats the structural-equation model as a directed network, with the value matrix of the network given by the structural coefficients of the model. Paths are traced by powering the value matrix, and effects are determined by summing the matrix powers, yielding the results given here after some algebraic manipulation.

nearly identical with the corresponding direct effects (obtained by regressing Y_5 on X_2 , Y_3 , and Y_4 , but not X_1), suggesting that the restriction $\gamma_{51} = 0$ is reasonable.

A good deal might be said about Table 4.10, but we wish to use the results in this table to illustrate just two important points. First, the fact that a prior variable has negligible direct effects does not necessarily mean that it is causally unimportant. Father's education (X_1), for example, has no direct effect on Y_5 , but it has nontrivial indirect effects. Likewise, the indirect effects of father's occupation (X_2) on Y_5 exceed the direct effects. Second, and in contrast to the first point, the fact that two variables have a strong empirical association does not imply that one exerts a strong causal impact on the other. For example, the implied slope relating Y_5 to first job (Y_4) is quite large, but nearly half of this association is spurious; indeed, in this case, the assumption that the disturbances ϵ_7 and ϵ_8 are uncorrelated is questionable (as we have noted), and even the direct-effect estimate B_{54} is probably inflated.

These points are not without consequence: Suppose that a naive investigator approaches the Blau and Duncan stratification data by regressing Y_5 on X_1 , X_2 , Y_3 , and Y_4 , as shown in the final column of Table 4.10. In light of the small (negative!) coefficient for X_1 , he or she might conclude that father's education has no impact on son's eventual occupational status. Because Y_3 and Y_4 intervene between X_1 and Y_5 , however, such a conclusion would be misleading:

TABLE 4.9. Effect Analysis for the Blau and Duncan Model

S^*_{YX}	X_1	X_2	S^*_{YY}	Y_3	Y_4	Y_5
Y_3	0.454 ^a	0.439 ^a	Y_3	1.001 ^a		
Y_4	0.315	0.417 ^a	Y_4	0.539 ^a	1.000 ^a	
Y_5	0.327	0.405 ^a	Y_5	0.596 ^a	0.541 ^a	1.001 ^a
D_{YX}	X_1	X_2	D_{YY}	Y_3	Y_4	Y_5
Y_3	0.310	0.279	Y_3			
Y_4	0.0	0.224	Y_4	0.440		
Y_5	0.0	0.115	Y_5	0.394	0.281	
I_{YX}	X_1	X_2	I_{YY}	Y_3	Y_4	Y_5
Y_3	0.0	0.0	Y_3			
Y_4	0.136	0.123	Y_4	0.0		
Y_5	0.161	0.207	Y_5	0.124	0.0	
N_{YX}	X_1	X_2	N_{YY}	Y_3	Y_4	Y_5
Y_3	0.144	0.160	Y_3			
Y_4	0.179	0.070	Y_4	0.098		
Y_5	0.166	0.083	Y_5	0.078	0.260	

Source: Fox (1980: Table 1).

^aNecessarily equal to observed value (within rounding error).

TABLE 4.10. Effect Analysis for Relationships of Prior Variables With Occupation in 1962 (Y_5), Blau and Duncan Model

Prior Variable j	Implied Slope M_{5j}^*	Direct Effect D_{5j}	Indirect Effect I_{5j}	Total Effect E_{5j}	Noncausal N_{5j}	Multiple Regression Slope B_{5j}^*
X_1 : Father's education	0.327	0.0	0.161	0.161	0.166	-0.014
X_2 : Father's occupation	0.405	0.115	0.207	0.322	0.083	0.121
Y_3 : Education	0.596	0.394	0.124	0.518	0.078	0.398
Y_4 : First job	0.541	0.281	0.0	0.281	0.260	0.281

Source: Adapted from Fox (1980: Table 2).

Indirect effects transmitted through Y_3 and Y_4 are ignored. In contrast, the reduction in the relationship between Y_4 and Y_5 when X_1 , X_2 , and Y_3 are controlled is properly interpreted as reflecting spurious sources of association. Proper interpretation, therefore, depends crucially upon an explicit or implicit causal model.

PROBLEMS

- 4.15. Perform a path analysis for (a) the standardized Duncan, Featherman, and Duncan model estimated in Problem 4.13; and (b) the unstandardized Lincoln model estimated in Problem 4.14.
- 4.16. In light of the material presented in this section, what are the risks for causal interpretation of controlling for "all relevant factors" in examining the relationship between two variables?

4.6. LATENT VARIABLES IN STRUCTURAL-EQUATION MODELS

The recent literature on latent variables in structural-equation models represents a confluence of work in several disciplines (see Goldberger, 1971, 1972; Griliches, 1977): in economics, on measurement errors in structural-equation and regression models; in psychology, on factor analysis and test theory; in biology, on path analysis; and in sociology, on constructs and their indicators. Sociologists also deserve credit for many of the applications of latent-variable models, and for an interest in integrating the several streams of work.

Latent variables (also called *unobserved variables*, *true variables*, *factors*, and *constructs*) arise for several related reasons: (1) Our measurements, even of relatively straightforward quantities, such as income and education, are imper-

fect. That is, observed variables generally have a measurement-error component. If this component is relatively large, it may be important to take it explicitly into account when we estimate structural relations, as will be shown in this section. (2) A variable appearing in a structural-equation model may be an abstract construct, such as racial prejudice, which is not directly observable. The construct may, however, have an observable effect, or *indicator*, and, indeed, it is often the case that multiple indicators are available. In general, however, no indicator is perfect; that is, each contains a measurement-error component. (3) Just as a construct may have observable indicators without itself being observable, likewise a construct may have one or more observable causes. Racial prejudice, for example, may be affected by education.

Social scientists frequently employ multiple indicators to construct composite scales prior to undertaking model building. While this strategy is reasonable, there are advantages to combining the processes of scale and model construction. First, a structural model may contribute to the definition of better scales, employing multiple indicators in a more efficient fashion. Second, taking measurement error into account explicitly may improve our estimates of structural parameters. As we shall see, however, multiple indicators and measurement errors cannot be included in a structural-equation model in a haphazard fashion. To build identified models, we have to make careful specifications that incorporate strong assumptions about the behavior of measurement errors. Although this is often a difficult undertaking, the alternative of ignoring errors in measurement can distort our findings.

The construction of structural-equation models with latent variables is a complex subject that we shall only be able to take up briefly here. Among the growing literature in this area, Duncan (1975: Ch. 9–10) presents a clear introductory treatment; a number of important papers appear in volumes edited by Goldberger and Duncan (1973) and by Aigner and Goldberger (1977), each of which contains extensive bibliographies. The LISREL computer program manual (Jöreskog and Sörbom, 1978, 1981) is also a valuable source (see Section 4.6.2).

4.6.1. Consequences of Random Measurement Error

In this section, we trace the consequences of random measurement error, prior to introducing a general model that accommodates latent variables in the next section. We determine the implications of measurement error for our usual estimators of structural parameters, and show how measurement error may be explicitly taken into account in the process of estimation. We proceed by examining three simple structural-equation models. Additional examples of this type may be found in Duncan (1975: Chapters 9–10), to which the exposition in this section is indebted.

Consider first the model shown in Figure 4.13. The notational conventions employed in this diagram anticipate the usage that we shall adopt in Section 4.6.2, and require some comment. The X 's and Y 's, as before in this chapter,

represent directly observed variables in mean-deviation form. In the present model, the X 's are exogenous variables, assumed to be measured without error, and the Y 's are fallible indicators of latent endogenous variables. The latent endogenous variables are symbolized by η 's. ζ 's are structural disturbances, while ε 's represent measurement errors in the endogenous indicators. Covariances are represented by σ 's. (Later on (in Section 4.6.2), we shall introduce additional notation for covariances among unobserved variables.)

The model in Figure 4.13 may conveniently be divided into two parts: First, there is the structural submodel:

$$\eta_5 = \gamma_{51}X_1 + \beta_{56}\eta_6 + \zeta_7 \quad (4.52)$$

$$\eta_6 = \gamma_{62}X_2 + \beta_{65}\eta_5 + \zeta_8$$

We make the usual distributional assumptions about the structural disturbances, and scale the latent variables so that they have zero expectations. Second, there is the measurement submodel:

$$Y_3 = \eta_5 + \varepsilon_9$$

$$Y_4 = \eta_6 + \varepsilon_{10}$$

We assume that the measurement errors, ε_9 and ε_{10} , are "well behaved," that is, each ε has an expectation of zero and is independent of all other variables in

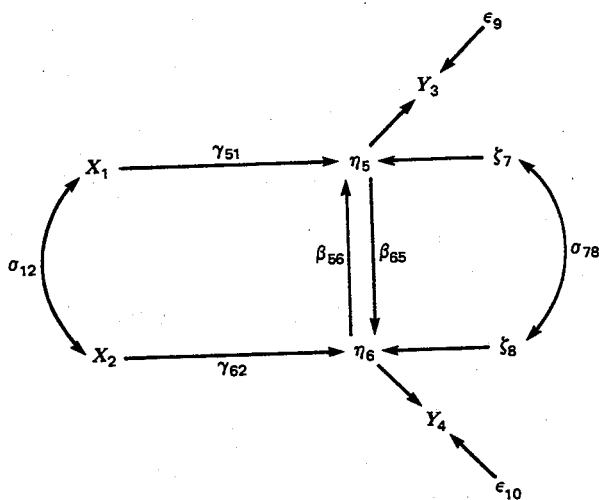


FIGURE 4.13. A structural-equation model in which the endogenous variables are measured with error.

the system, save the indicator with which it is associated. In other words, the measurement errors are "random," not systematic.¹⁵

One way of approaching latent-variable models is to eliminate the latent variables from the structural equations, substituting for these variables by employing the relations specified in the measurement submodel. This approach is particularly fruitful for our current purpose of determining the consequences of ignoring measurement error in the endogenous variables. We shall work with the first structural equation in (4.52), exploiting the symmetry of the model. Substituting for η_5 and η_6 , we get

$$(Y_3 - \epsilon_9) = \gamma_{51}X_1 + \beta_{56}(Y_4 - \epsilon_{10}) + \zeta_7$$

which we may rewrite as

$$Y_3 = \gamma_{51}X_1 + \beta_{56}Y_4 + \zeta'_7 \tag{4.53}$$

where $\zeta'_7 \equiv \zeta_7 + \epsilon_9 - \beta_{56}\epsilon_{10}$. In effect, we merge the measurement errors of the endogenous indicators with the structural disturbance. Equation (4.53) may be estimated in the usual manner, because our instrumental variables, X_1 and X_2 , are uncorrelated not only with the structural disturbance ζ_7 but also with the measurement errors ϵ_9 and ϵ_{10} , and consequently with the composite disturbance ζ'_7 .

This result is general: In a nonrecursive model, with no disturbance-covariance restrictions, we may safely ignore random measurement error in the endogenous variables for purposes of estimating structural parameters of the model, so long as the exogenous variables are measured without error.

For a contrasting example, we shall examine the model in Figure 4.14. The structural equations for this model are

$$Y_4 = \gamma_{46}\xi_6 + \gamma_{42}X_2 + \zeta_7 \tag{4.54}$$

$$Y_5 = \gamma_{53}X_3 + \beta_{54}Y_4 + \zeta_8$$

and the measurement-submodel equation is

$$X_1 = \xi_6 + \delta_9 \tag{4.55}$$

Here, X_1 is a fallible indicator, with random measurement error δ_9 , of the latent exogenous variable ξ_6 . Proceeding as before, let us substitute for ξ_6 in the first structural equation:

$$\begin{aligned} Y_4 &= \gamma_{46}(X_1 - \delta_9) + \gamma_{42}X_2 + \zeta_7 \\ &= \gamma_{46}X_1 + \gamma_{42}X_2 + \zeta'_7 \end{aligned} \tag{4.56}$$

¹⁵Under certain circumstances, it is possible to estimate models specifying measurement errors that are correlated with each other. The general model developed in Section 4.6.2 permits such a specification.

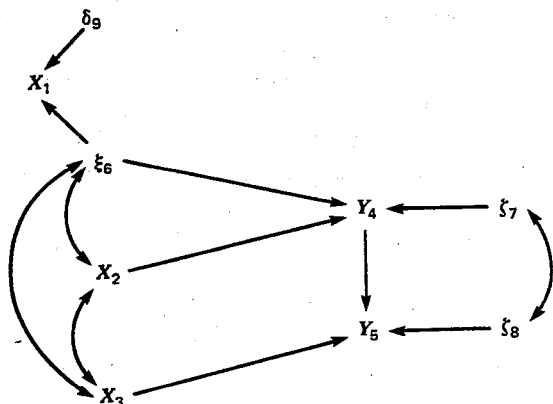


FIGURE 4.14. A structural-equation model in which an exogenous variable is measured with error.

where $\zeta'_7 \equiv \zeta_7 - \gamma_{46}\delta_9$.

Multiplying equation (4.56) through by X_1 and X_2 , and taking expectations, we get

$$\sigma_{14} = \gamma_{46}\sigma_{11} + \gamma_{42}\sigma_{12} - \gamma_{46}\sigma_{99} \quad (4.57)$$

$$\sigma_{24} = \gamma_{46}\sigma_{12} + \gamma_{42}\sigma_{22}$$

Note that $E(X_1\zeta'_7) = -\gamma_{46}\sigma_{99}$ because X_1 is the sum of ξ_6 and δ_9 , both of which are uncorrelated with ζ_7 , and because $\sigma_{19} = \sigma_{99}$ due to the uncorrelation of δ_9 and ξ_6 . Similarly, $E(X_2\zeta'_7) = 0$ because X_2 is uncorrelated with both ζ_7 and δ_9 . Equations (4.57) may be solved for the structural parameters γ_{46} and γ_{42} ; we find it convenient to write the solution in the following manner (after Duncan, 1975: 120):

$$\gamma_{46} = \frac{\sigma_{14}\sigma_{22} - \sigma_{12}\sigma_{24}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2 - \sigma_{99}\sigma_{22}} \quad (4.58)$$

$$\gamma_{42} = \frac{\sigma_{11}\sigma_{24} - \sigma_{12}\sigma_{14}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} - \frac{\gamma_{46}\sigma_{12}\sigma_{99}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}$$

Imagine, now, that we make the mistake of treating X_1 as if it were measured without error. In this instance, we would attempt to estimate the first structural equation by OLS, since both independent variables, X_1 and X_2 , are "exogenous." In other words, we would apply X_1 and X_2 as instrumental variables, wrongly setting σ_{99} in equations (4.58) to zero. The population analog of the OLS estimator of γ_{46} , then, is given by

$$\gamma'_{46} = \frac{\sigma_{14}\sigma_{22} - \sigma_{12}\sigma_{24}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}$$

The denominator of γ_{46} is necessarily positive. Since the factor missing from the denominator of γ'_{46} (i.e., $-\sigma_{99}\sigma_{22}$) is necessarily negative, γ'_{46} is biased towards zero. In general, ignoring measurement error in an exogenous independent variable tends to attenuate its coefficient (although the combined effects of measurement errors in several exogenous independent variables are indeterminate for the individual coefficients—see the next paragraph).

The population analog of the OLS estimator of γ_{42} is

$$\gamma'_{42} = \frac{\sigma_{11}\sigma_{24} - \sigma_{12}\sigma_{14}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}$$

Put alternatively, $\gamma'_{42} = \gamma_{42} + \text{bias}$, where the sign of the bias term depends upon the signs of γ_{46} and σ_{12} . In general, if an exogenous independent variable is measured with error, ignoring that error will have an indeterminate effect on the coefficients of other independent variables in the structural equation.

In the present example, X_1 should not be used as an IV for estimating the first structural equation. Note, however, that X_1 is eligible as an IV for purposes of estimating the second structural equation. Because δ_9 , the measurement-error component of X_1 , is uncorrelated with all other variables in the system, the covariance of X_1 with any *other* variable is the same as the covariance of ξ_6 with that variable; for example, multiplying equation (4.55) by Y_4 and taking expectations produces

$$\sigma_{14} = \sigma_{46} + \sigma_{49} = \sigma_{46}$$

Although we cannot legitimately estimate the first structural equation in model (4.54) by OLS, this equation is identified because both X_2 and X_3 are available as IVs: Each of these variables is uncorrelated with ξ'_7 in equation (4.56). Note that the first structural equation would ordinarily be overidentified, but the overidentifying restriction has been “consumed” by the measurement error in X_1 . It is often the case that strategically placed overidentifying structural restrictions may serve to identify a model with a measurement-error component.

In the current example, it is also possible to estimate the *measurement-error variance*, σ_{99} , along with the *true-score variance*, σ_{66} . Squaring the measurement-submodel equation (4.55), and taking expectations, we obtain

$$\sigma_{11} = \sigma_{66} + \sigma_{99} \tag{4.59}$$

due to the uncorrelation of ξ_6 and δ_9 . From equation (4.57) we have

$$\sigma_{99} = \frac{\gamma_{46}\sigma_{11} + \gamma_{42}\sigma_{12} - \sigma_{14}}{\gamma_{46}} \tag{4.60}$$

The σ 's on the right-hand side of equation (4.60) may be estimated directly from sample data, and we have already seen that we can obtain IV estimates of

the structural parameters γ_{46} and γ_{42} . With an estimate of σ_{99} in hand, we may estimate σ_{66} by subtraction, using equation (4.59). For this model, then, the overidentifying restriction on the first structural equation not only permits us to estimate the structural parameters of the model, but also serves to identify the measurement submodel. Incidentally, notice that in the previous example (Figure 4.13), the variances of the measurement errors cannot be separated from the variances of the structural disturbances, rendering the measurement submodel (but not the structural parameters) underidentified.

The model shown in Figure 4.15 provides us with a third and final example, which incorporates multiple indicators X_1 and X_2 of a latent exogenous variable ξ_6 . As before, we assume that the random measurement errors δ_9 and δ_{10} are well behaved—that is, have zero expectations, are uncorrelated with each other, and are uncorrelated with the other variables in the model except the indicators with which they are associated. The model consists of two structural equations,

$$Y_4 = \gamma_{46}\xi_6 + \beta_{45}Y_5 + \zeta_7$$

$$Y_5 = \gamma_{53}X_3 + \beta_{54}Y_4 + \zeta_8$$

and two measurement equations,

$$X_1 = \xi_6 + \delta_9 \tag{4.61}$$

$$X_2 = \lambda\xi_6 + \delta_{10}$$

The coefficient for ξ_6 in the measurement equation for X_1 is implicitly one, while that in the equation for X_2 is an unknown parameter λ . By arbitrarily fixing the coefficient of the latent variable in one of the measurement-submodel equations, we in effect express the latent variable in the metric (i.e., units of measurement) of the corresponding indicator. This choice of scale is an essentially arbitrary normalization rule: We could equally well fix the coeffi-

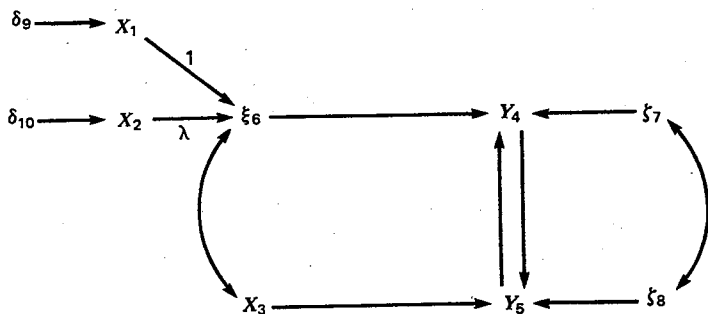


FIGURE 4.15. A structural-equation model with multiple indicators of an exogenous variable.

cient of ξ_6 to one in the measurement equation for X_2 , in which case the coefficient for ξ_6 in the equation for X_1 becomes $1/\lambda$. Suppose, for example, that ξ_6 is socio-economic status, X_1 is education measured in years, and X_2 is income measured in dollars; according to equation (4.61), the latent variable ξ_6 is measured in years, and λ converts years to dollars.¹⁶

We may analyze the present model, as before, by substituting for the unobserved variable in the first structural equation. Now, however, because we have two indicators of ξ_6 , we may substitute in two different ways:

$$\begin{aligned} Y_4 &= \gamma_{46}(X_1 - \delta_9) + \beta_{45}Y_5 + \zeta_7 \\ &= \gamma_{46}X_1 + \beta_{45}Y_5 + \zeta_7' \end{aligned} \quad (4.62)$$

where $\zeta_7' \equiv \zeta_7 - \gamma_{46}\delta_9$; and

$$\begin{aligned} Y_4 &= \gamma_{46}\left(\frac{X_2}{\lambda} - \frac{\delta_{10}}{\lambda}\right) + \beta_{45}Y_5 + \zeta_7 \\ &= \frac{\gamma_{46}}{\lambda}X_2 + \beta_{45}Y_5 + \zeta_7'' \end{aligned} \quad (4.63)$$

where $\zeta_7'' \equiv \zeta_7 - (\gamma_{46}/\lambda)\delta_{10}$.

Let us now multiply each of equations (4.62) and (4.63) by X_3 and take expectations, obtaining

$$\sigma_{34} = \gamma_{46}\sigma_{13} + \beta_{45}\sigma_{35}$$

$$\sigma_{34} = \frac{\gamma_{46}}{\lambda}\sigma_{23} + \beta_{45}\sigma_{35}$$

which we may solve for $\lambda = \sigma_{23}/\sigma_{13}$. This solution is not unique, however, for we may obtain alternative expressions for λ by taking expectations of equations (4.62) and (4.63) with Y_4 and Y_5 (rather than X_3); for example, for Y_4 ,

$$\sigma_{44} = \gamma_{46}\sigma_{14} + \beta_{45}\sigma_{45} + \sigma_{47}$$

$$\sigma_{44} = \frac{\gamma_{46}}{\lambda}\sigma_{24} + \beta_{45}\sigma_{45} + \sigma_{47}$$

which yields $\lambda = \sigma_{24}/\sigma_{14}$. Likewise, applying Y_5 produces $\lambda = \sigma_{25}/\sigma_{15}$. Y_4 and Y_5 serve to obtain expressions for λ because, although they are endogenous, they are uncorrelated with the measurement errors δ_9 and δ_{10} . If the model is correctly specified, then $\sigma_{23}/\sigma_{13} = \sigma_{24}/\sigma_{14} = \sigma_{25}/\sigma_{15}$; in the sample, however, we cannot expect these relations to hold precisely, and thus the parameter λ is overidentified.

¹⁶Although this example serves to illustrate clearly the arbitrary choice of scale for a latent variable, and though this type of specification for socio-economic status has been employed in applications, it is more sensible in this instance to conceive of the latent variable status as an *effect* of education and income rather than as a *cause* of these observable variables.

With knowledge of λ , we may proceed to determine the parameters in the first structural equation. Applying X_2 to equation (4.62), we get

$$\sigma_{24} = \gamma_{46}\sigma_{12} + \beta_{45}\sigma_{25} \quad (4.64)$$

Even though X_2 is measured with error, it is uncorrelated with δ_9 . Similarly, applying X_1 to equation (4.63) produces

$$\sigma_{14} = \frac{\gamma_{46}}{\lambda}\sigma_{12} + \beta_{45}\sigma_{15} \quad (4.65)$$

Because we have already determined λ , equations (4.64) and (4.65) may be solved for the structural parameters γ_{46} and β_{45} . Since alternative estimates of λ are available, and since our estimates of γ_{46} and β_{45} depend upon which value we use, these structural parameters are also overidentified.

The following points concerning this last example are noteworthy: (1) If there were only one fallible indicator of ξ_6 , the measurement submodel and the structural submodel would both be underidentified. (2) If ξ_6 were observed directly and measured without error, the structural submodel would be just identified. (3) The presence of two fallible indicators of ξ_6 in the example serves to overidentify both the measurement submodel and the structural submodel.

4.6.2. Specification and Estimation of LISREL Models

LISREL, an acronym for *linear structural relations*, refers both to a general structural-equation model with latent variables and multiple indicators (Jöreskog, 1973; Jöreskog and Sörbom, 1977), and to a computer program (Jöreskog and Sörbom, 1978, 1981) that provides full-information maximum-likelihood estimates for this model.¹⁷ The highly general nature of the LISREL model permits a variety of specifications; for example, by specifying that all indicators are measured without error, and by establishing a one-to-one correspondence between indicators and latent variables, the LISREL model becomes an ordinary structural-equation model, for which the LISREL program computes the usual FIML estimates.

In this section, we consider the form of the general LISREL model and sketch its estimation; we also examine an illustrative application. Although we should generally establish the identification of a LISREL model prior to estimating its parameters, the identification of models with latent variables is a sufficiently involved topic to warrant separate treatment; we therefore take up this subject in the next section.

¹⁷The version of LISREL described in this section is LISREL IV (Jöreskog and Sörbom, 1978). The newer LISREL V (Jöreskog and Sörbom, 1981) incorporates a slightly different structural model. We employ the LISREL IV model because its format is more similar to that of the general structural-equation model presented in Section 4.1.

Because the LISREL model is complex, we adopt the notation established by Jöreskog and his colleagues, even though this notation conflicts to a degree with the conventions employed in this text. To do otherwise would be to invite confusion, for the reader will surely have occasion to consult other sources, such as the LISREL program manual. The symbols employed in the LISREL model are summarized in Table 4.11.

The LISREL model may conveniently be divided into two parts: (1) a *structural submodel*, specifying relations among latent variables; and (2) a *measurement submodel*, which links latent variables to their observed indicators. We shall examine each submodel in turn.

The structural submodel has the familiar form of a structural-equation model:

$$\underset{(m \times m)}{\mathbf{B}} \underset{(m \times 1)}{\boldsymbol{\eta}_i} = \underset{(m \times n)}{\boldsymbol{\Gamma}} \underset{(n \times 1)}{\boldsymbol{\xi}_i} + \underset{(m \times 1)}{\boldsymbol{\zeta}_i}$$

$\boldsymbol{\eta}_i$, $\boldsymbol{\xi}_i$, and $\boldsymbol{\zeta}_i$ are, consecutively, vectors of latent endogenous variables, latent exogenous variables, and structural disturbances, for the i th of N observations. Henceforth, we shall generally suppress the subscript i for observation. These, and indeed all, variables in the LISREL model are expressed as deviations from their expectations. \mathbf{B} and $\boldsymbol{\Gamma}$ are matrices of structural parameters. Note that in the LISREL model, the exogenous variables appear on the right-hand side of the structural equations. The covariance matrix of the exogenous variables is given by $\underset{(n \times n)}{\boldsymbol{\Phi}}$, and the covariances of the structural disturbances by $\underset{(m \times m)}{\boldsymbol{\Psi}}$. We assume that \mathbf{B} is nonsingular, that $\boldsymbol{\xi}_i \sim N_n(\mathbf{0}, \boldsymbol{\Phi})$, that $\boldsymbol{\zeta}_i \sim N_m(\mathbf{0}, \boldsymbol{\Psi})$, that $\boldsymbol{\zeta}_i$ is independent of $\boldsymbol{\xi}_i$, and that the observations are independent.¹⁸

The reduced form of the LISREL model is

$$\begin{aligned} \boldsymbol{\eta} &= \mathbf{B}^{-1}\boldsymbol{\Gamma}\boldsymbol{\xi} + \mathbf{B}^{-1}\boldsymbol{\zeta} \\ &= \mathbf{D}\boldsymbol{\xi} + \mathbf{B}^{-1}\boldsymbol{\zeta} \end{aligned}$$

where $\underset{(m \times n)}{\mathbf{D}} \equiv \mathbf{B}^{-1}\boldsymbol{\Gamma}$ is the matrix of reduced-form parameters. The covariance matrix $\underset{(m \times m)}{\mathbf{C}}$ of the latent endogenous variables is, therefore,

$$\begin{aligned} \mathbf{C} &= E(\boldsymbol{\eta}\boldsymbol{\eta}') = E\left[(\mathbf{D}\boldsymbol{\xi} + \mathbf{B}^{-1}\boldsymbol{\zeta})(\mathbf{D}\boldsymbol{\xi} + \mathbf{B}^{-1}\boldsymbol{\zeta})'\right] \\ &= \mathbf{D}E(\boldsymbol{\xi}\boldsymbol{\xi}')\mathbf{D}' + \mathbf{D}E(\boldsymbol{\xi}\boldsymbol{\zeta}')(\mathbf{B}^{-1})' + \mathbf{B}^{-1}E(\boldsymbol{\zeta}\boldsymbol{\xi}')\mathbf{D}' + \mathbf{B}^{-1}E(\boldsymbol{\zeta}\boldsymbol{\zeta}')(\mathbf{B}^{-1})' \\ &= \mathbf{D}\boldsymbol{\Phi}\mathbf{D}' + \mathbf{B}^{-1}\boldsymbol{\Psi}(\mathbf{B}^{-1})' \end{aligned} \tag{4.66}$$

¹⁸For models in which the exogenous variables are measured without error (see below), we need not make assumptions about their distribution—as was the case for the observed-variable structural-equation models considered earlier in the chapter.

TABLE 4.11. Notation for the LISREL Model

Symbol	Meaning
N	Number of observations
m	Number of latent endogenous variables
n	Number of latent exogenous variables
p	Number of indicators of latent endogenous variables
q	Number of indicators of latent exogenous variables
η	Vector of latent endogenous variables
$(m \times 1)$	
ξ	Vector of latent exogenous variables
$(n \times 1)$	
ζ	Vector of structural disturbances (errors in equations)
$(m \times 1)$	
B	Structural parameter matrix relating latent endogenous variables
$(m \times m)$	
Γ	Structural parameter matrix relating latent endogenous to latent exogenous variables
$(m \times n)$	
D	Reduced-form coefficient matrix
$(m \times n)$	
y	Vector of indicators of latent endogenous variables
$(p \times 1)$	
x	Vector of indicators of latent exogenous variables
$(q \times 1)$	
ϵ	Vector of measurement errors in endogenous indicators
$(p \times 1)$	
δ	Vector of measurement errors in exogenous indicators
$(q \times 1)$	
Λ_y	Coefficient matrices relating indicators to latent variables
$(p \times m)$	
Λ_x	
$(q \times n)$	
Φ	Matrix of covariances among latent exogenous variables
$(n \times n)$	
Ψ	Matrix of covariances among structural disturbances
$(m \times m)$	
Θ_ϵ	Matrices of covariances among measurement errors
$(p \times p)$	
Θ_δ	
$(q \times q)$	
Σ	Matrix of covariances among observed (indicator) variables
$(p+q \times p+q)$	
C	Matrix of covariances among latent endogenous variables
$(m \times m)$	

Because the metric of the latent variables is frequently not substantively meaningful, the researcher may wish to standardize the latent variables to unit variance. Let us define diagonal matrices of standard deviations

$$A_{\eta} \equiv (\text{diag } C)^{1/2}$$

(m × m)

$$A_{\xi} \equiv (\text{diag } \Phi)^{1/2}$$

(n × n)

Then $\eta^* \equiv A_{\eta}^{-1}\eta$ and $\xi^* \equiv A_{\xi}^{-1}\xi$ are vectors of standardized latent variables. Each of the coefficient and covariance matrices previously defined may be simply adjusted to provide a standardized solution; for example

$$B^* = A_{\eta}^{-1}BA_{\eta}$$

$$\Gamma^* = A_{\eta}^{-1}\Gamma A_{\xi}$$

$$\Psi^* = A_{\eta}^{-1}\Psi A_{\eta}^{-1}$$

The LISREL measurement submodel consists of two matrix equations:

$$y_i = \Lambda_y \eta_i + \epsilon_i$$

(p × 1) (p × m) (m × 1) (p × 1)

$$x_i = \Lambda_x \xi_i + \delta_i$$

(q × 1) (q × n) (n × 1) (q × 1)

Here, y_i and x_i are vectors of indicators of the latent endogenous and exogenous variables, respectively; ϵ_i and δ_i are vectors of measurement-error variables, one for each indicator; and Λ_y and Λ_x are matrices of regression coefficients relating the indicators to the latent variables. In general, each column of Λ_y and Λ_x contains one unit entry, to fix the metric of the corresponding latent variable, as explained in the previous section [see equation (4.61)]. (Alternatively, the variance of a latent exogenous variable may be fixed, as may the variance of the disturbance associated with a latent endogenous variable.) It is, moreover, frequently the case that each row of Λ_y and Λ_x has but one nonzero entry: Each observed variable is an indicator of just one latent variable. In certain cases, however, it may make sense to treat an observed variable as the effect of more than one latent variable, and LISREL accommodates such a specification.

The covariances of the measurement errors appear in Θ_{ϵ} and Θ_{δ} .

(p × p) (q × q)

These matrices are not necessarily diagonal; that is, measurement errors may be correlated. Unless they are specified carefully and frugally, however, corre-

lated measurement errors are likely to underidentify a model. The measurement errors ε_i and δ_i are assumed to have zero expectations; to be independent of each other and of η_i , ξ_i , and ζ_i ; and to be multivariately normally distributed: $\varepsilon_i \sim N_p(\mathbf{0}, \Theta_\varepsilon)$, $\delta_i \sim N_q(\mathbf{0}, \Theta_\delta)$.

An illustrative LISREL model (from Jöreskog and Sörbom, 1978) for the Duncan, Haller, and Portes peer-influences data is shown in Figure 4.16. In this model, there are multiple indicators for the latent endogenous variables, but the exogenous variables are assumed to be measured without error. We have, then, the following LISREL specifications for the exogenous indicators:

$$\begin{matrix} \mathbf{x} & = & \boldsymbol{\xi} & \text{(that is, } \Lambda_x = \mathbf{I}_6, \boldsymbol{\delta} = \mathbf{0} \text{); and consequently } \boldsymbol{\Phi} = \Sigma_{xx}, \\ (6 \times 1) & & (6 \times 1) & & (6 \times 1) & & (6 \times 6) \end{matrix}$$

$$\Theta_\delta = \mathbf{0} \quad (\text{where } \Sigma_{xx} \text{ is the covariance matrix of the exogenous variables}).$$

For the endogenous indicators,

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \lambda_{21}^y & 0 \\ 0 & 1 \\ 0 & \lambda_{42}^y \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}$$

$$\Theta_\varepsilon = \text{diag}(\theta_{11}^\varepsilon, \theta_{22}^\varepsilon, \theta_{33}^\varepsilon, \theta_{44}^\varepsilon)$$

(4 × 4)

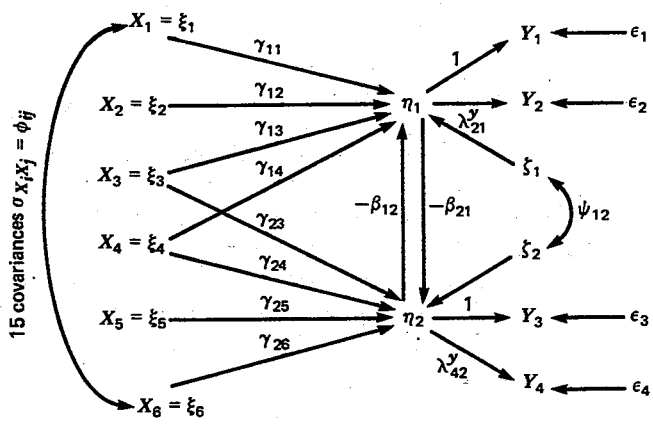


FIGURE 4.16. A LISREL model with latent endogenous variables for the Duncan, Haller, and Portes peer-influences data. X_1 , respondent's parental aspiration; X_2 , respondent's family SES; X_3 , respondent's intelligence; X_4 , best friend's intelligence; X_5 , best friend's family SES; X_6 , best friend's parental aspiration; Y_1 , respondent's occupational aspiration; Y_2 , respondent's educational aspiration; Y_3 , best friend's occupational aspiration; Y_4 , best friend's educational aspiration; η_1 , respondent's general level of aspiration; η_2 , best friend's general level of aspiration. (Source: Adapted with permission from Karl G. Jöreskog and Dag Sörbom, 1978, copyright National Educational Resources, Inc., 1978, 1981; and Duncan, Haller, and Portes, 1968, see Table 4.2.)

Θ_ϵ is diagonal because the measurement errors in the endogenous indicators are specified to be uncorrelated. The structural submodel is given by

$$\begin{pmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & 0 & 0 \\ 0 & 0 & \gamma_{23} & \gamma_{24} & \gamma_{25} & \gamma_{26} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \\ \xi_6 \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix}$$

From the LISREL structural and measurement submodels we may derive expressions for the covariances of the indicators. It is in this manner that a link is established between the parameters of the model and observable quantities. Let Σ represent the covariance matrix for the indicators,

$$\Sigma_{(p+q \times p+q)} = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix}$$

$(p \times p)$ $(p \times q)$
 $(q \times p)$ $(q \times q)$

We have, for example,

$$\begin{aligned} \Sigma_{yy} &= E(\mathbf{yy}') = E[(\Lambda_y \eta + \epsilon)(\Lambda_y \eta + \epsilon)'] \\ &= \Lambda_y E(\eta \eta') \Lambda_y' + \Lambda_y E(\eta \epsilon') + E(\epsilon \eta') \Lambda_y' + E(\epsilon \epsilon') \\ &= \Lambda_y \mathbf{C} \Lambda_y' + \Theta_\epsilon \end{aligned}$$

Then, using equation (4.66):

$$\Sigma_{yy} = \Lambda_y [\mathbf{B}^{-1} \Gamma \Phi \Gamma' (\mathbf{B}^{-1})' + \mathbf{B}^{-1} \Psi (\mathbf{B}^{-1})'] \Lambda_y' + \Theta_\epsilon \tag{4.67}$$

We may determine the other components of Σ in a similar manner, obtaining (see Problem 4.19)

$$\begin{aligned} \Sigma_{yx} &= \Lambda_y \mathbf{B}^{-1} \Gamma \Phi \Lambda_x' \\ \Sigma_{xy} &= \Sigma_{yx}' = \Lambda_x \Phi \Gamma' (\mathbf{B}^{-1})' \Lambda_y' \\ \Sigma_{xx} &= \Lambda_x \Phi \Lambda_x' + \Theta_\delta \end{aligned} \tag{4.68}$$

The covariances of the observed variables, therefore, are functions (albeit complex functions) of the parameters of the LISREL structural and measure-

ment model. In addition, the assumptions of the model insure that the observed variables follow a multivariate normal distribution:

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} \sim N_{p+q}(\mathbf{0}, \Sigma)$$

Thus, the joint probability density for the observed variables in a sample of size N is given by

$$p(\mathbf{y}_1, \mathbf{x}_1; \dots; \mathbf{y}_N, \mathbf{x}_N) = \frac{1}{(2\pi)^{N(p+q)/2} |\Sigma|^{N/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^N \begin{pmatrix} \mathbf{y}_i \\ \mathbf{x}_i \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} \mathbf{y}_i \\ \mathbf{x}_i \end{pmatrix}\right)$$

The logarithm of the likelihood function may be written (Problem 4.19)

$$\begin{aligned} \log L(\mathbf{B}, \Gamma, \Phi, \Psi, \Theta_e, \Theta_s, \Lambda_y, \Lambda_x) \\ = -\frac{N(p+q)}{2} \log(2\pi) - \frac{N}{2} [\log|\Sigma| + \text{trace}(\mathbf{S}\Sigma^{-1})] \end{aligned} \quad (4.69)$$

where $\mathbf{S} = (1/N)[\mathbf{Y}, \mathbf{X}][\mathbf{Y}, \mathbf{X}]'$ is the sample covariance matrix for the indicators. Recall that Σ is a function of the model parameters (equations (4.67) and (4.68)). Equation (4.69) tells us, in essence, that the likelihood is large when the covariance matrix $\hat{\Sigma}$ implied by the model is similar to the observed sample covariance matrix \mathbf{S} .

In estimating the model, it is necessary to take account of the prior constraints on the parameters. Some of these constraints are normalizations: Certain parameters are prespecified to be one. Other constraints are exclusions: Certain parameters are prespecified to be zero. (The LISREL computer program also permits equality constraints, where two or more parameters are prespecified to be equal to each other.) The maximum-likelihood estimators of \mathbf{B} , Γ , and the other parameters maximize L subject to these prior constraints. As in the case of FIML estimation of observed-variable models, the log likelihood (4.69) for the LISREL model must be maximized numerically. Estimated asymptotic standard errors for estimators of all "free" (i.e., unconstrained) parameters may be obtained in the usual manner from the inverse of the information matrix, which the LISREL program computes.

The example model shown in Figure 4.16 was estimated for standardized indicators. Correlations among these indicators were given earlier in Table 4.2. The standardized maximum-likelihood solution appears in equations (4.70). (Note that although λ_{11}^y and λ_{32}^y originally were fixed to one, these values change along with the other entries of Λ_y when the model is standardized.)

Standard errors for free-parameter estimates are shown in parentheses beneath the estimates.

$$\hat{\Lambda}_y^* = \begin{pmatrix} 0.7667 & 0 \\ 0.8148 & 0 \\ (0.0691) & \\ 0 & 0.8299 \\ 0 & 0.7716 \\ & (0.0583) \end{pmatrix}$$

$$\hat{\Theta}_\varepsilon^* = \text{diag} \left(\begin{matrix} 0.4121 & 0.3361 & 0.3112 & 0.4046 \\ (0.0512) & (0.0521) & (0.0459) & (0.0462) \end{matrix} \right)$$

$$\hat{B}^* = \begin{pmatrix} 1 & -0.1994 \\ & (0.1027) \\ -0.2176 & 1 \\ (0.1103) & \end{pmatrix} \tag{4.70}$$

$$\hat{\Gamma}^* = \begin{pmatrix} 0.2103 & 0.3256 & 0.2848 & 0.0937 & 0 & 0 \\ (0.0506) & (0.0574) & (0.0576) & (0.0648) & & \\ 0 & 0 & 0.0746 & 0.2758 & 0.4205 & 0.1922 \\ & & (0.0624) & (0.0532) & (0.0546) & (0.0468) \end{pmatrix}$$

$$\hat{\Psi}^* = \begin{pmatrix} 0.4780 \\ (0.0786) \\ -0.0355 & 0.3830 \\ (0.0804) & (0.0648) \end{pmatrix}$$

These results seem to be substantively acceptable: All of the structural and measurement parameter estimates are of the anticipated sign, and all but $\hat{\gamma}_{14}$ and $\hat{\gamma}_{23}$ are statistically significant (at or beyond approximately the 2.5 percent level, one-tail). Since γ_{14} and γ_{23} represent the effects of the other boy's SES on each boy's aspirations, it is reasonable that these coefficients be small; indeed, in an earlier model for the peer-influences data (Figure 4.1), we set these parameters to zero *a priori*. When we estimated that earlier model, we found a large negative correlation between the structural disturbances. Notice that for the present LISREL model, the covariance between the disturbances, $\hat{\psi}_{12}$, is reassuringly close to zero, although a positive value would be even more reasonable. Apparently, the introduction of additional exogenous variables (parental aspirations, X_1 and X_6), and the recognition of measurement error in the indicators of the endogenous variables (made possible by multiple indicators) has had a salutary effect on our estimates (see Gillespie and Fox, 1980).

4.6.3. Identification of Models With Latent Variables

The identification of structural-equation models with latent variables is a complex problem, and one that does not yet have a straightforward, general

solution. Some progress, however, has been made. Geraci (1977), for example, discusses general conditions for identification of single-indicator models in which (some) exogenous variables are measured with error. Further discussion of the identification problem for latent-variable models may be found in Wiley (1973).

It is always possible to demonstrate the identification of a model, if indeed it is identified, by showing that each of its parameters may be expressed in at least one way as a function of covariances of observed variables. To do this, we may use methods such as those employed in Section 4.6.1 and later in the present section. This process is a tedious one, and if we fail to demonstrate the identifiability of a model, we often cannot be certain that it is not our imagination that has failed rather than the model.

There are, however, necessary conditions for identification, which may show us that a model is underidentified; and there are sufficient conditions which insure that a model is identified. What is missing is a condition, analogous to the rank condition for nonrecursive observed-variable models, that is both necessary and sufficient. Through the application of known necessary conditions and sufficient conditions, we hope to avoid having to identify a model on an individual basis.

In practice, we may proceed to attempt to estimate a model without having established its identification. An underidentified model produces a singular information matrix, because the infinity of solutions implied by underidentification is reflected in a flat likelihood function at the maximum.¹⁹ The LISREL computer program calculates the information matrix, and therefore is generally able to detect an attempt to estimate an underidentified model.

A global necessary condition for identification is that there be no more free parameters to estimate than there are unique covariances among the observed variables in the system. Put another way, the number of unknowns in the estimating equations cannot exceed the number of known quantities. There are $(p + q + 1)(p + q)/2$ unique observable variances and covariances, a number that is greatly exceeded by the number of entries in \mathbf{B} , Γ , Φ , Ψ , Θ_ϵ , Θ_δ , Λ_x , and Λ_y , even after normalizations are taken into account. Many prior restrictions, therefore, are needed to identify a LISREL model. Some of these restrictions may derive trivially from the particular application, as when a variable is specified to be measured without error.

The model in Figure 4.16, for example, has 40 free parameters: λ_{21}^y , λ_{42}^y , β_{12} , β_{21} , eight γ 's, ψ_{11} , ψ_{12} , ψ_{22} , four θ^{ϵ} 's, and $6(7)/2 = 21$ ϕ 's (which necessarily equal the corresponding covariances among observed exogenous variables). There are $(4 + 6 + 1)(4 + 6)/2 = 55$ unique entries in Σ ; if the model is identified, therefore (and it is), there are 15 overidentifying restrictions. This counting rule does not guarantee the identification of a model: Although restrictions are sufficiently numerous, they may be injudiciously located.

¹⁹In this respect, underidentification is analogous to perfect collinearity, which also leads to underdetermined estimating equations.

It is often possible to establish the identification of a model by treating the measurement and structural submodels separately. We then seek to show (1) that the restrictions on the measurement submodel are sufficient to identify all covariances among latent variables, and (2) that, given these covariances, the structural submodel is identified. For a nonrecursive structure, the conditional identification of the structural submodel may be assessed by the rank condition. One useful rule of thumb (i.e., sufficient condition) is that the measurement submodel is identified if (but not only if): (1) there are at least two unique indicators for each latent variable, or if there is just one indicator, it is measured without error; and (2) measurement errors are uncorrelated. It must be stressed that the separation of the measurement and structural submodels, and hence the scope of this rule of thumb, are restrictive: As we saw in the previous section, measurement and structural submodels may contribute to each other's identification; likewise, models with single fallible indicators or with correlated measurement errors may frequently be identified.

We demonstrate first that the variances and covariance of two latent variables are identified if each has two fallible indicators, and if measurement errors are independent. We refer to the model diagrammed in Figure 4.17. Here, there are 10 observable covariances among the four indicators. There are nine unknown parameters: λ_{21}^x , λ_{42}^x , ϕ_{11} , ϕ_{12} , ϕ_{22} , θ_{11}^δ , θ_{22}^δ , θ_{33}^δ , and θ_{44}^δ . To show that the parameters may be expressed in terms of observed covariances, we expand these covariances by the expectation method:

$$(a) \quad \sigma_{11} = \phi_{11} + \theta_{11}^\delta$$

$$(b) \quad \sigma_{22} = \phi_{11}\lambda_{21}^x + \theta_{22}^\delta$$

$$(c) \quad \sigma_{33} = \phi_{22} + \theta_{33}^\delta$$

$$(d) \quad \sigma_{44} = \phi_{22}\lambda_{42}^x + \theta_{44}^\delta$$

$$(e) \quad \sigma_{12} = \phi_{11}\lambda_{21}^x$$

$$(f) \quad \sigma_{34} = \phi_{22}\lambda_{42}^x$$

$$(g) \quad \sigma_{13} = \phi_{11}\phi_{12}\phi_{22}$$

$$(h) \quad \sigma_{14} = \phi_{11}\phi_{12}\phi_{22}\lambda_{42}^x$$

$$(i) \quad \sigma_{23} = \lambda_{21}^x\phi_{11}\phi_{12}\phi_{22}$$

$$(j) \quad \sigma_{24} = \lambda_{21}^x\phi_{11}\phi_{12}\phi_{22}\lambda_{42}^x$$

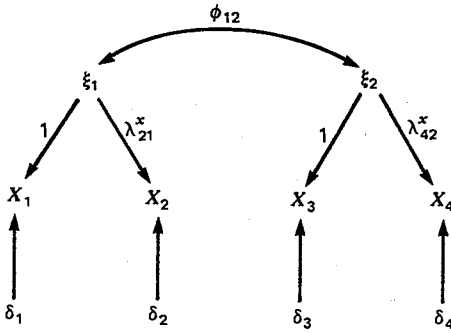


FIGURE 4.17. Two indicators for each of two latent variables.

Solving first for the λ 's, we get

$$\lambda_{21}^x = \frac{(i)}{(g)} = \frac{\sigma_{23}}{\sigma_{13}} = \frac{(j)}{(h)} = \frac{\sigma_{24}}{\sigma_{14}}$$

$$\lambda_{42}^x = \frac{(h)}{(g)} = \frac{\sigma_{14}}{\sigma_{13}} = \frac{(j)}{(i)} = \frac{\sigma_{24}}{\sigma_{23}}$$

showing that these parameters are overidentified. With knowledge of the λ 's, we derive expressions for the other parameters by successive substitution:

$$(e') \quad \phi_{11} = \frac{\sigma_{12}}{\lambda_{21}^x}$$

$$(f') \quad \phi_{22} = \frac{\sigma_{34}}{\lambda_{42}^x}$$

$$(g') \quad \phi_{12} = \frac{\sigma_{13}}{\phi_{11}\phi_{22}}$$

$$(a') \quad \theta_{11}^\delta = \sigma_{11} - \phi_{11}$$

$$(b') \quad \theta_{22}^\delta = \sigma_{22} - \phi_{11}\lambda_{21}^x$$

$$(c') \quad \theta_{33}^\delta = \sigma_{33} - \phi_{22}$$

$$(d') \quad \theta_{44}^\delta = \sigma_{44} - \phi_{22}\lambda_{42}^x$$

Next, let us analyze the model shown in Figure 4.18, where one latent variable is measured without error and the other has two indicators with uncorrelated measurement errors. There are $3(4)/2 = 6$ observed covariances

and an equal number of free parameters: λ_{32}^x , ϕ_{11} , ϕ_{12} , ϕ_{22} , θ_{22}^δ , and θ_{33}^δ . Expressing the observed covariances as functions of the parameters, we get:

- (a) $\sigma_{11} = \phi_{11}$
- (b) $\sigma_{22} = \phi_{22} + \theta_{22}^\delta$
- (c) $\sigma_{33} = \lambda_{32}^x \phi_{22} + \theta_{33}^\delta$
- (d) $\sigma_{12} = \phi_{11} \phi_{12} \phi_{22}$
- (e) $\sigma_{13} = \phi_{11} \phi_{12} \phi_{22} \lambda_{32}^x$
- (f) $\sigma_{23} = \phi_{22} \lambda_{32}^x$

Solving for the parameters, which are just identified, completes the demonstration:

$$\lambda_{32} = \frac{(e)}{(d)} = \frac{\sigma_{13}}{\sigma_{12}}$$

$$(a') \quad \phi_{11} = \sigma_{11}$$

$$(f') \quad \phi_{22} = \frac{\sigma_{23}}{\lambda_{32}^x}$$

$$(d') \quad \phi_{12} = \frac{\sigma_{12}}{\phi_{11} \phi_{22}}$$

$$(b') \quad \theta_{22}^\delta = \sigma_{22} - \phi_{22}$$

$$(c') \quad \theta_{33}^\delta = \sigma_{33} - \lambda_{32}^x \phi_{22}$$

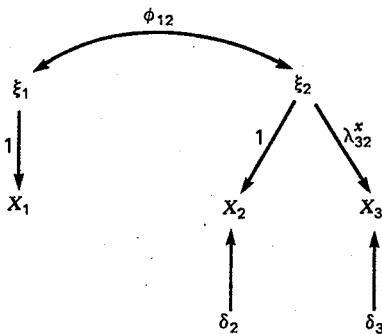


FIGURE 4.18. Two indicators for one latent variable and another variable measured without error.

PROBLEMS

4.17. The model in Figure 4.19 is adapted from work by Bielby, Hauser, and Featherman (1977) on response errors in models of the stratification process. Alternative measures collected on different occasions were available for each of the latent variables in the model:

ξ_1 Father's occupational status

ξ_2 Father's education

ξ_3 Parents' income

η_1 Respondent's education

η_2 Occupational status of the respondent's first job

η_3 Current occupational status

(a) Comment on the specification of the model, paying attention to the measurement-model assumptions as well as to the structural model.

(b) Show that the model is identified.

(c) Using the correlations for a sample of 578 nonblack males in the civilian labor force, given in Table 4.12, estimate the parameters of the model. Is it important to take measurement error into account in estimating this model?

4.18. The data in Table 4.13 were compiled by Inverarity (1976) as part of a study of the relationship between "mechanical solidarity" and "repres-

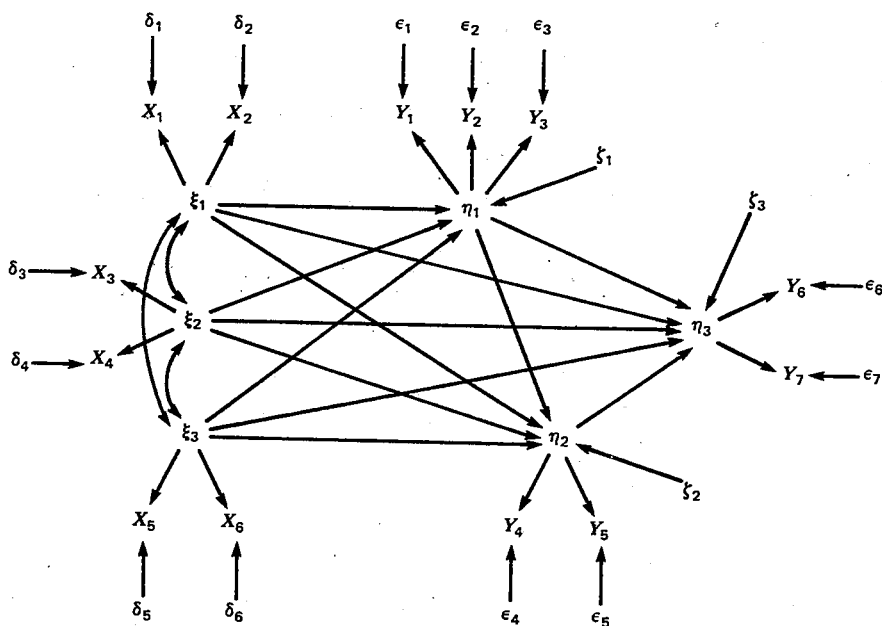


FIGURE 4.19. Bielby, Hauser, and Featherman's stratification model.

sive justice.” Both concepts, and the hypothesized link between them, are due to the 19th-century French sociologist, Emile Durkheim, who distinguished between two forms of social solidarity: mechanical solidarity based on social similarity, and organic solidarity based on a detailed social division of labor. Inverarity examines the incidence of lynching in the period 1889–1896 for 59 Louisiana parishes, relating this measure of repressive justice to characteristics of the parishes. He argues that proportion black (X_1) and religious homogeneity (X_3) should exert a positive influence on mechanical solidarity among whites, and that urbanization (X_2 , coded one for parishes with some urban population, and zero otherwise) should have a negative impact. Inverarity predicts that the number of lynchings in a parish (Y_3) is directly related to the level of mechanical solidarity (η_1) and to the size of the black population (X_4 , in thousands) in the parish. The proportion of the Democratic Party vote in the 1892 presidential election (Y_1) and in the 1896 gubernatorial election (Y_2) are taken as indicators of mechanical solidarity. The model in Figure 4.20 was specified by Bagozzi (1977) for Inverarity’s data. (Also see Wasserman, 1977; Pope and Ragin, 1977; Bohrnstedt, 1977; and Inverarity, 1977).

TABLE 4.12. Bielby, Hauser, and Featherman’s Stratification Data

	X_1	X_2	X_3	X_4	X_5	X_6	Y_1	Y_2
X_1	1.000							
X_2	0.869	1.000						
X_3	0.585	0.589	1.000					
X_4	0.597	0.599	0.939	1.000				
X_5	0.422	0.437	0.477	0.467	1.000			
X_6	0.426	0.450	0.436	0.478	0.913	1.000		
Y_1	0.428	0.430	0.448	0.445	0.426	0.439	1.000	
Y_2	0.445	0.443	0.483	0.492	0.485	0.502	0.838	1.000
Y_3	0.419	0.419	0.467	0.467	0.486	0.501	0.801	0.921
Y_4	0.398	0.410	0.290	0.300	0.370	0.358	0.581	0.644
Y_5	0.409	0.409	0.325	0.322	0.363	0.348	0.578	0.642
Y_6	0.340	0.369	0.280	0.284	0.291	0.296	0.504	0.563
Y_7	0.364	0.390	0.291	0.308	0.307	0.301	0.519	0.603

	Y_3	Y_4	Y_5	Y_6	Y_7
Y_3	1.000				
Y_4	0.637	1.000			
Y_5	0.631	0.847	1.000		
Y_6	0.534	0.585	0.599	1.000	
Y_7	0.566	0.618	0.620	0.797	1.000

Source: Bielby, Hauser, and Featherman, in Aigner and Goldberger, eds., *Latent Variables in Socio-economic Models*, © Elsevier North-Holland, Inc., 1977. Reprinted with permission from Bielby, Hauser, and Featherman (1977: Table 2).

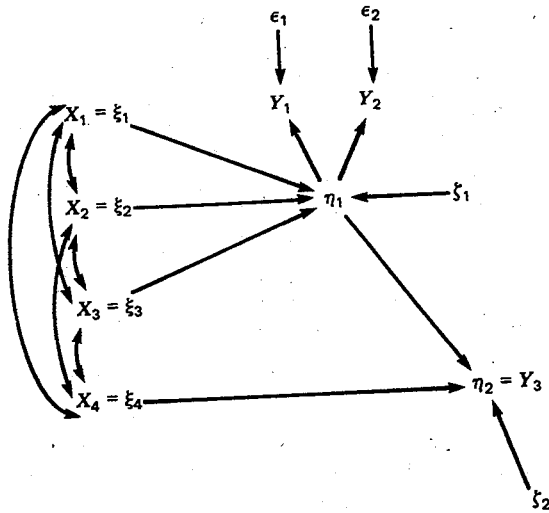


FIGURE 4.20. Bagozzi's model for Inverarity's data on lynchings.

- (a) Comment on the specification of Bagozzi's model. Does it adequately capture Inverarity's argument?
 - (b) Show that the model is identified. What happens to the identification status of the model if Y_3 is specified to be measured with error (that is, if Y_3 is taken as an imperfect indicator of repressive justice)?
 - (c) Estimate the model using the covariances in Table 4.13.
- 4.19. (a) Derive equations (4.68) for Σ_{yx} , Σ_{xy} , and Σ_{xx} from the LISREL model.
- (b) Derive the log likelihood for the LISREL model [equation (4.69)].

TABLE 4.13. Covariances for Inverarity's Lynching Data

	X_1	X_2	X_3	X_4	Y_1	Y_2	Y_3
X_1	0.04537						
X_2	-0.01029	0.19272					
X_3	0.00403	-0.01566	0.00672				
X_4	0.53791	1.94136	-0.20256	121.61678			
Y_1	0.01449	0.00903	0.00424	0.25706	0.04410		
Y_2	0.03166	-0.00583	0.00334	0.58084	0.02385	0.05244	
Y_3	0.09991	0.11866	-0.00183	7.32943	0.08448	0.07756	2.52810
Mean	0.508	0.254	0.902	11.031	0.766	0.589	1.41

Source: Adapted with permission from Inverarity (1976: Table 2).

4.7. EVALUATION OF STRUCTURAL-EQUATION MODELS

As is general with statistical models, having fit a structural-equation model to data, it is desirable to determine, to the extent possible, whether the model adequately represents the data. Certain checks on the adequacy of the model are implicit in the process of interpretation and testing, as when we ask whether an estimated coefficient assumes a reasonable value, or is statistically distinguishable from zero.

Some procedures for analysis of residuals, presented in Section 3.2 for single-equation linear models, may be extended to structural-equation models (see Belsley, Kuh, and Welsch, 1980: 266–269). For recursive models, the application of these procedures is straightforward, since a recursive model is simply a collection of related regression equations, each estimated by OLS. For nonrecursive models, estimated say by 2SLS, endogenous independent variables are in general correlated with structural residuals. In examining residual plots for nonlinearity, therefore, we must discount visually whatever linear relation is present. An alternative is to use second-stage regression residuals for certain residual analyses, for example, in the detection of outliers.

Two topics related to the question of model quality are dealt with at greater length in this section. First, we develop measures of fit for structural-equation models; and second, we assess the adequacy of overidentifying restrictions.

We should not conclude, however, that all assumptions underlying a structural-equation model may be examined in light of the data. As pointed out in Section 4.5, we cannot in general expect data to mediate issues of causal priority. Nor can we generally assess the crucial assumption of independence between exogenous variables and disturbances: Indeed, in a just-identified model, exogenous variables and structural residuals are perfectly uncorrelated in the sample, much as the independent variables are uncorrelated with the residuals in OLS regression. In an overidentified equation, however, structural residuals may have nonzero correlations with excluded exogenous variables. If these correlations are substantial, we should question the adequacy of the overidentifying restrictions, a topic pursued in Section 4.7.2.

4.7.1. Indices of Fit for Structural-Equation Models

For endogenous variables measured in a meaningful metric, the estimated standard error, S_{E_j} , provides an interpretable measure of fit, as in linear regression analysis. Since recursive models are fit by OLS, there is a multiple correlation coefficient for each structural equation; these R^2 values may be interpreted in the usual manner.

Although several R^2 analogs have been proposed for structural equations in nonrecursive models, these statistics do not have certain properties that we

associate with the multiple correlation coefficient. In OLS regression

$$R^2 = \frac{\text{TSS} - \text{RSS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\frac{\text{RSS}}{n}}{\frac{\text{TSS}}{n}} = 1 - \frac{S_E^2}{S_Y^2}$$

Though the statistic $1 - (S_E^2/S_Y^2)$ may be computed for a structural equation in a nonrecursive model, and though this statistic seems sensible, Basmann (1962) has shown that it is unbounded below. Likewise, in OLS regression, $R = r_{\hat{Y}Y}$. In a nonrecursive structural-equation model, the correlation between observed and fitted endogenous-variable values may be negative (Basmann, 1962). Nevertheless, statistics such as $1 - (S_E^2/S_Y^2)$ and $r_{\hat{Y}Y}$ are frequently reported because of their simplicity and intuitive appeal.

Since the reduced form may be estimated consistently by OLS regression, we may justifiably report the multiple correlation for each reduced-form equation (i.e., for each endogenous variable). Hooper (1959) has extended this approach, formulating a correlational index for nonrecursive structural-equation models that assesses the degree of joint dependence of the endogenous on the exogenous variables. This dependence is expressed in the reduced-form equation

$$\mathbf{Y} = \mathbf{X} \Pi' + \Delta$$

$(n \times q) \quad (n \times m) \quad (m \times q) \quad (n \times q)$

Hooper estimates the reduced-form parameters by OLS regression:

$$\mathbf{P}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

(An alternative would be to obtain $\mathbf{P} = -\mathbf{B}^{-1}\mathbf{C}$, but in an overidentified model this procedure would change the ultimate interpretation of Hooper's correlational measure.) Fitted endogenous-variable values are then given by

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{P}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

The multivariate analog of total variation (TSS) is the sum-of-squares-and-products matrix $\mathbf{Y}'\mathbf{Y}$; similarly, $\hat{\mathbf{Y}}'\hat{\mathbf{Y}}$ is the analog of "explained" variation (RegrSS). $\hat{\mathbf{Y}}'\hat{\mathbf{Y}}(\mathbf{Y}'\mathbf{Y})^{-1}$, therefore, is a multivariate version of explained "divided by" total variation. Hooper's *trace correlation* statistic is defined as

$$\bar{R}^2 = \frac{1}{q} \text{trace}[\hat{\mathbf{Y}}'\hat{\mathbf{Y}}(\mathbf{Y}'\mathbf{Y})^{-1}] \quad (4.71)$$

Hooper demonstrates that \bar{R}^2 is the mean squared canonical correlation²⁰

²⁰Canonical correlations assess the strength of linear dependencies between two sets of variables—here, the exogenous and endogenous variables in a structural-equation model. For details of canonical-correlation theory, see Morrison (1976: 259–263).

between the sets of endogenous and exogenous variables. Notice that when $q = 1$, equation (4.71) specializes to the usual R^2 statistic.

For the Duncan, Haller, and Portes peer-influences model in Figure 4.1, $\bar{R}^2 = .230$. Roughly, then, 23 percent of the joint variation of Y_4 and Y_6 is accounted for by their relation to X_1 , X_2 , X_3 , and X_4 . The individual R^2 's for the two reduced-form regressions are .264 and .319.

4.7.2. Overidentification Tests

We have mentioned at several points in this chapter that an overidentified structural-equation model may be inconsistent with the observed data. One descriptive method for tracing the consequences of overidentification is to calculate model-implied covariances among observed variables, comparing these to sample covariances computed directly from the data. Implied covariances may be calculated by equations (4.51), or for LISREL models, by the sample analogs of equations (4.67) and (4.68). Implied covariances (correlations) for the standardized Blau and Duncan model were shown in Table 4.9. It is clear that this model closely reproduces covariances among observed variables.²¹

A formal test of overidentifying restrictions may be constructed by the likelihood-ratio principle. We develop this test for the LISREL model, because of that model's generality.²²

The log likelihood for the LISREL model is given by equation (4.69). Suppose that a model with r free parameters is overidentified and has a likelihood of L_0 . If the overidentifying restrictions are removed, S and the model-implied estimate of Σ become identical, yielding log likelihood

$$\log L_1 = -\frac{N(p+q)}{2} \log(2\pi) - \frac{N}{2} (\log|S| + p + q)$$

The number of free parameters in the just-identified model is equal to the number of observed covariances: $(p+q)(p+q+1)/2$. The likelihood-ratio test statistic for the overidentifying restrictions is therefore

$$\begin{aligned} G_0^2 &= -2(\log L_0 - \log L_1) \\ &= N [\log|\hat{\Sigma}| + \text{trace}(S\hat{\Sigma}^{-1}) - \log|S| - (p+q)] \end{aligned} \quad (4.72)$$

Here $\hat{\Sigma}$ is the estimate of Σ implied by the maximum-likelihood estimates of

²¹Even an overidentified recursive model necessarily reproduces certain covariances precisely. See Fox (1980) for further discussion of this point. For the Blau and Duncan model, in fact, only S_{14}^* and S_{15}^* may depart from the corresponding S_{ij} 's.

²²Overidentification tests are also available for single-equation methods such as 2SLS. See, for example, Fox (1979a).

the model parameters. As a likelihood-ratio statistic, G_0^2 is asymptotically distributed as χ^2 with $[(p + q)(p + q + 1)/2] - r$ degrees of freedom (the degree of overidentification of the model). Notice that G_0^2 will be small when the reproduced covariances $\hat{\Sigma}$ are similar to their directly observed counterparts S . Since observed-variable structural-equation models estimated by FIML are special cases of LISREL estimation, the test in equation (4.72) is applicable to these models as well.

We have noted that OLS and FIML are identical for recursive models. For these models, the overidentification test statistic in equation (4.72) specializes to (Land, 1973):

$$G_0^2 = n \sum_{j=1}^q \log \frac{S_{E_j}^2}{S_{E_j^*}^2}$$

where n is the sample size, q is the number of equations in the model, $S_{E_j}^2$ is the estimated residual variance from the j th structural equation, and $S_{E_j^*}^2$ is the estimated residual variance for the j th equation respecified so that no prior variables are excluded. If the $S_{E_j}^2$ are appreciably larger than the $S_{E_j^*}^2$, then we should suspect that prior variables have been falsely excluded; when this is the case, the test statistic will be large. The degrees of freedom for G_0^2 are equal to the number of excluded prior variables in all structural equations of the model.

Overidentification test statistics for the models discussed in this chapter are shown in Table 4.14. Notice that the extremely large sample for the Blau and Duncan study results in a statistically significant overidentification test even though the model fits the data very closely: Given a large enough sample, virtually any overidentified model may be rejected. (Indeed, for the Blau and Duncan model, the difference between the calculated G_0^2 and zero may be the

TABLE 4.14. Overidentification Tests

Model	n	G_0^2	df	p
Duncan et al. Nonrecursive (Figure 4.1)	329	2.81	2	.25
Blau and Duncan Recursive (Figure 4.2)	20,700	30.9	2	$\ll .001$
Duncan et al. Block-Recursive (Figure 4.4)	329	3.81	2	.15
Duncan et al. LISREL (Figure 4.16)	329	26.70	15	.03

result of rounding errors.) The significant overidentification test for the LISREL model fit to the peer-influences data is more distressing, yet even here we should be loath to reject the model on this basis alone.

A statistically significant overidentification test may motivate model respecification, perhaps by removing one or more overidentifying restrictions, perhaps by more drastic reformulation. The data may be helpful in suggesting which overidentifying restrictions to remove, but we must be careful always to guide our model-building activity by substantive criteria.

The magnitude of differences between model-implied and directly calculated sample covariances is an uncertain guide to model respecification (see Costner and Schoenberg, 1973; Sörbom, 1975). Sörbom (1975) has suggested examining the partial derivatives of the likelihood function with respect to the *fixed* (i.e., constrained) parameters. These derivatives are available as by-products of the model-fitting procedure. When the likelihood is maximized, its partial derivatives with respect to the *free* parameters are, of course, zero; a steep gradient with respect to a *fixed* parameter, however, indicates that the likelihood might be substantially increased if that parameter were permitted to take on a different value.

PROBLEMS

- 4.20. Compute trace correlations and reduced-form correlations for the non-recursive models estimated in Problems 4.11 and 4.12.
- 4.21. Calculate model-implied covariances and likelihood-ratio overidentification tests for the structural-equation models fit in Problems 4.12, 4.13, 4.14, 4.17, and 4.18. In each case, if the model appears to be inadequate, consider how it might be respecified.