10

Multiple Indicators

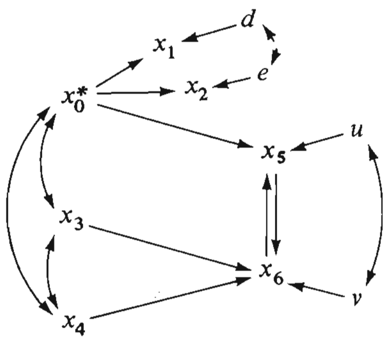
Terminology to designate unobserved variables tends to vary by discipline. In psychometrics, where one fallible score is at stake, the unobserved counterpart is called a "true score"; errors of measurement give rise to "unreliability"; and using an estimate of reliability to correct the estimates of structural coefficients is termed "correcting for attenuation." Where there are several (in practice, four or more) fallible measurements for each of one or more unobserved variables, the latter are conceptualized as "factors," and the theory of factor analysis is brought into play.

In economics, recognition of measurement error (random or systematic) is often symbolized by referring to observed variables as "proxy" variables, acknowledging that they are not identical with the (unobserved) variables discussed in economic theory. The examples in Chapter 9 disclose some ways in which estimates of structural coefficients may be distorted by naively replacing the unobserved variable in an equation by its "proxy." Economists are also sensitized to the possibility of hypothetical constructs like "permanent income," for which actual measurements of income (even if free of observational error) afford only a proxy, owing to the random variability introduced into actual money receipts from time to time by transitory causes.

In sociology, this generic problem often is referred to as the problem of validity of "indicators." Perhaps there is no direct measure of such a concept as "social cohesion," for example, but one theory of the phenomenon would suggest that the suicide rate is an "indicator" of social cohesion. There are many reasons, however, why the relationship between concept and indicator is contingent and loose rather than fixed and exact. In view of these reasons (we will not spell them out here), it is sometimes felt that the investigator is better off with multiple indicators, on the theory (we surmise) that there is then a chance that an error in one indicator may be offset by a compensating error in another.

Here, we can only illustrate a few of the issues that arise when an investigator resolves to take seriously the problem of indicator validity. Our premise is that there is no *general* solution to this problem, but that it has to be attacked on its merits in connection with each substantive model in which indicator variables appear. The reader is warned at the outset that a major purpose of this chapter is to provide motivation for further study. Hence, we repeatedly note problems in estimation, optimal solutions to which require resort to methods beyond the scope of this volume.

Let us modify slightly the last example in Chapter 9:



The new feature is that there are now two distinct indicators (or fallible measurements) of the unobserved variable. Hence, we have the two equations,

$$x_1 = x_0^* + d$$

$$x_2 = a x_0^* + e$$

Note that one of these equations must include a "scale factor" a since, in general, we do not require that x_1 and x_2 be measured on the same scale. (Suicide rate and per capita contributions to social welfare agencies might be two such noncommensurable indicators of "social cohesion"; but the reader must henceforth supply his own illustrative names for variables.) It is simply a matter of convenience which of the two indicators we designate as x_1 (without a scale factor). The consequence of this choice is that we thereby establish an implicit metric for the unobserved variable. We require the errors, d and e, to be "well behaved," so that they are uncorrelated with x_0^* and also with observed variables $x_3, \ldots, x_6$. As usual, predetermined variables are uncorrelated with structural disturbances. Thus, x_0^*, x_3, and x_4 are all uncorrelated with both u and v (although the latter may well have a nonzero covariance). These specifications suffice to insure that both x_1 and x_2 are uncorrelated with structural disturbances u and v.

Exercise. *Verify this, if it is not already "obvious."*

We leave open for later discussion the question of whether $E(de) = 0$, and the diagram shows a dashed curve connecting the errors d and e. The structural equations are

$$x_5 = b_{50} x_0^* + b_{56} x_6 + u$$

$$x_6 = b_{63} x_3 + b_{64} x_4 + b_{65} x_5 + v$$

We note that the status of each of these equations with respect to identification is made problematic by the presence in the model of the unobserved variable and the indicator variables.

A general heuristic device for studying models containing unobserved variables is to "solve out" such variables, thereby deriving new equations. These are then examined in regard to identification and problems of estimation. Let us substitute $x_0^* = x_1 - d$ (from the x_1-equation) for x_0^* in the x_2-equation:

$$x_2 = a x_1 + e' \quad \text{where} \quad e' = e - ad$$

We see that four of the variables in the model, $x_3, \ldots, x_6$, are eligible as instruments, since each is uncorrelated with both d and e and, therefore, with e'. At first glance it may seem odd that the list of instrumental variables includes the two endogenous variables, x_5 and x_6. A closer look reveals that the model is, in effect, block recursive, with the x_5- and x_6-equations comprising one block, the x_1- and x_2-equations the other. With this embarrassment of riches, the rewritten x_2-equation—or, precisely, its one coefficient, a—is clearly overidentified. We are tempted to proceed at once to estimation, resolving the problem of an excess number of instrumental variables in a manner reminiscent of 2SLS. That is, we would regress x_1 on x_3, x_4, x_5, and x_6 to calculate $\hat{x}_1$ and then estimate a from the OLS regression of x_2 on $\hat{x}_1$. OLS is legitimate at this stage, since $E(\hat{x}_1 e') \cong 0$, given that $\hat{x}_1$ is a linear combination of $x_3, \ldots, x_6$, each of which is uncorrelated with e'.

Before beginning this calculation, however, we should note that our new equation can also be written,

$$x_1 = \frac{1}{a} x_2 + d'$$

where

$$d' = d - \frac{e}{a}$$

Again, we have an overidentified equation and, again, it would perhaps seem that a two-stage regression procedure would provide a straightforward approach. Moreover, it hardly matters whether we estimate a or $1/a$; knowledge of one is equivalent to knowledge of the other, in principle. The unfortunate fact is, however, that such a procedure is sensitive to the choice of normalization rule. That is, if the estimate $\hat{a}$ is obtained from the solved-out x_2-equation and $\widehat{1/a}$ from the solved-out x_1-equation (using two-stage regression with the same list of instrumental variables), then $1/\hat{a} \neq \widehat{1/a}$ in general. Moreover, the regression results provide no basis for preferring one of these estimates to the other or for averaging or otherwise reconciling them. The awkward situation is that this estimation problem does not yield nicely to the seemingly straightforward two-stage regression approach. If the investigator finds this too disconcerting, he must resort to more advanced methods—too advanced to fall within the scope of this book.

[Not only the two-stage regression procedure mentioned here, but also 2SLS itself, in the standard simultaneous-equation context, is sensitive to the normalization rule. We did not mention this in Chapter 7, on the assumption that the investigator will have resolved the question of normalization before proceeding to estimation. Fisher (reprinted in Blalock, 1971) remarks: "This is not an unreasonable assumption, as such normalization rules are generally present in model building, each variable of the model being naturally associated with that particular endogenous variable which is determined by the decision-makers whose behavior is represented by the equation [page 264]." Whether or not Fisher's view is plausible, there clearly is no "natural" normalization rule in the present instance.]

There is another instructive way to look at the overidentification of the x_1- and x_2-equations. Let us multiply through each of them by all the eligible instrumental variables; we find

$$\begin{aligned} \sigma_{13} &= \sigma^*_{03} & \sigma_{23} &= a\sigma^*_{03} \\ \sigma_{14} &= \sigma^*_{04} & \sigma_{24} &= a\sigma^*_{04} \\ \sigma_{15} &= \sigma^*_{05} & \sigma_{25} &= a\sigma^*_{05} \\ \sigma_{16} &= \sigma^*_{06} & \sigma_{26} &= a\sigma^*_{06} \end{aligned}$$

The availability of several distinct ways of calculating the parameter a means that overidentifying restrictions are implied in setting those solutions equal to each other:

$$a = \frac{\sigma_{23}}{\sigma_{13}} = \frac{\sigma_{24}}{\sigma_{14}} = \frac{\sigma_{25}}{\sigma_{15}} = \frac{\sigma_{26}}{\sigma_{16}}$$

Although these equalities must hold in the population (if the model is true), they can be expected to hold at best only approximately in the sample. But if there are marked departures from equality in the sample data (too large to attribute to sampling error), we have evidence of some sort of specification error. Later examples will suggest some ways in which this could come about. But it should be noted—you have presumably come to expect this kind of "bad news"—that rejection of the overidentifying restrictions (or even some particular subset of them) does not provide unambiguous evidence concerning the nature and location of the specification error. At best, this outcome provides clues to the specification error, and such clues must be interpreted with great care and caution (Costner & Schoenberg, 1973).

Let us turn our attention to the other two equations, which, after all, comprise the substance of the model. Looking first at the x_6-equation, which has three explanatory variables, we find that there are actually four variables available as instruments: $x_1, \ldots, x_4$, since each of them is uncorrelated with the disturbance v. (At this juncture, by the way, it does not matter whether $\sigma_{de} = 0$.) Hence, the x_6-equation is overidentified, and it seems reasonable to proceed with two-stage estimation, regressing x_5 on x_1, x_2, x_3, and x_4 to calculate $\hat{x}_5$ and then regressing x_6 on x_3, x_4, and $\hat{x}_5$. It is a comfort to know that at least this much of the model is relatively unproblematic. However, if there is any concern about the correctness of our specifications on the x_1- and x_2-equations, this had best be resolved first, because any respecification of these equations might affect the eligibility of these variables as instruments.

To estimate the x_5-equation, we will need to "solve out" the unobserved variable. For example, using the x_1-equation for this purpose we obtain

$$x_5 = b_{50}x_1 + b_{56}x_6 + u'$$

where

$$u' = u - b_{50}d$$

It appears that the equation is just identified, since only x_3 and x_4 are available as instrumental variables, if we wish to keep open the possibility that $\sigma_{de} \neq 0$. On the other hand, if $\sigma_{de} = 0$, then x_2 as well, though not x_1 (*Why?*), can serve as an instrument. Working with just x_3 and x_4, IV estimates of the structural coefficients may be obtained as solutions to

$$m_{35} = \hat{b}_{50}m_{13} + \hat{b}_{56}m_{36}$$

$$m_{45} = \hat{b}_{50}m_{14} + \hat{b}_{56}m_{46}$$

But our conclusion was too hasty, for we could equally well have used the x_2-equation to "solve out" x_0^* from the x_5-equation:

$$x_5 = (b_{50}/a)x_2 + b_{56}x_6 + u''$$

where

$$u'' = u - b_{50}e/a$$

and IV estimates would be

$$m_{35} = \widehat{(b_{50}/a)}m_{23} + \hat{b}_{56}m_{36}$$

$$m_{45} = \widehat{(b_{50}/a)}m_{24} + \hat{b}_{56}m_{46}$$

We would then obtain $\hat{b}_{50}$ making use of our previously obtained estimate, $\hat{a}$ (supposing that little problem to have been solved!). The two sets of IV estimates would not, in general, be the same, although they should be "approximately" so if the model is true. We must conclude that the x_5-equation is overidentified, but this comes about in such a way that two-stage regression does not provide a way to resolve the problem.

Indeed, the problem is not merely that of deciding which one of these two estimators to accept. We can "solve out" x_0^* from the x_5-equation using not only the x_1-equation or the x_2-equation, but any weighted combination of the two equations. Suppose W is a constant. We may write

$$Wx_1 = Wx_0^* + Wd$$

$$(1 - W)x_2 = (1 - W)ax_0^* + (1 - W)e$$

whence

$$x_0^* = \frac{Wx_1 + (1 - W)x_2 - Wd - (1 - W)e}{W + (1 - W)a}$$

This solution, for whatever value of W, may be used to "solve out" x_0^* from the x_5-equation. (If $W = 1$, we are using, in effect, the x_1-equation alone, as before; if $W = 0$, we are using the x_2-equation alone.) Once this is done—supposing a value of W and an estimate of a are available to substitute into the formula—we can proceed to IV estimation of the x_5-equation using x_3 and x_4 as instrumental variables.

But the calculation of W is our new stumbling block, if we insist on seeking a statistically efficient method of estimation, rather than one that merely tends to be unbiased in large samples. Clearly, W should be chosen in such a way that the sampling variance of our estimates, $\hat{b}_{50}$ and $\hat{b}_{56}$, is a minimum. The solution to this difficult problem is, again, beyond our scope. (Actually, this formulation of the issue, in terms of choosing an optimal value of W, is introduced only to illustrate the complexity of the estimation problem that arises with multiple indicators. We do not mean to suggest that calculation of W will be an explicit step in an efficient method of estimation. On the contrary, such a method will, in effect, accomplish the requisite weighting at the same time that it produces the appropriate estimate of the overidentified parameter, a. This same remark applies at later points in the chapter, where similar issues of "weighting" arise.)

One final aspect of this model to be considered arises in evaluating the variances and covariance of our indicator variables. We find, by the usual technique, that

$$\sigma_{11} = \sigma_{00}^{*} + \sigma_{dd}$$

$$\sigma_{22} = a^2\sigma_{00}^{*} + \sigma_{ee}$$

$$\sigma_{12} = a\sigma_{00}^{*} + \sigma_{de}$$

We see that if σ_{de} is specified to be zero, the parameters σ_{00}^{*}, σ_{dd}, and σ_{ee} are just identified; for in that event the three equations can be solved for the three unknowns. Assuming we have already estimated a, we could use the sample data to compute

$$\hat{\sigma}_{00}^{*} = \frac{\hat{\sigma}_{12}}{\hat{a}}$$

$$\hat{\sigma}_{dd} = \hat{\sigma}_{11} - \hat{\sigma}_{00}^{*}$$

$$\hat{\sigma}_{ee} = \hat{\sigma}_{22} - \hat{a}^2\hat{\sigma}_{00}^{*}$$

This operation is something of a side issue, in that it contributes nothing to estimation of coefficients in our two main equations. It may be of some use, however, in appraising our indicators; for we can now examine the values of $\hat{\sigma}_{dd}/\hat{\sigma}_{11}$ and $\hat{\sigma}_{ee}/\hat{\sigma}_{22}$ as indices of unreliability. In future research, if forced to choose between the indicators, we would presumably prefer the one with the lesser value of this index. But in using this criterion, we should want to be rather sure of the specification $\sigma_{de} = 0$, for there is no way to test it with data on our six observed variables; nor does the value of σ_{de} (if it is not zero) have any other direct consequence for our procedures.

Summing up, what has been gained by the use of the two indicators? (We seemed to get along perfectly well with only one in the otherwise similar model in Chapter 9.) First, in view of the overidentifying restrictions on the parameter a, we have an opportunity to test whether our indicators really are "well behaved," as the model assumes them to be. If the truth of the overidentifying restrictions is not called into question by the result of a suitable test, we are somewhat reassured on this score. If the outcome is otherwise, we know that one or both indicators are "contaminated" in a way that the model fails to take into account. We must reconsider the model or the indicators.

Second, assuming we continue to accept the overidentifying restrictions after testing them, we have a clue as to which indicator is more reliable, if that information is of any use, and if we are willing to make the strong assumption that the indicators are correlated only to the extent of their common dependence on the unobserved variable.

Third, again on the assumption that the overidentifying restrictions are acceptable, the sampling errors of our estimates of structural coefficients ordinarily will be smaller when using two (well-behaved) indicators, rather than either of them alone.

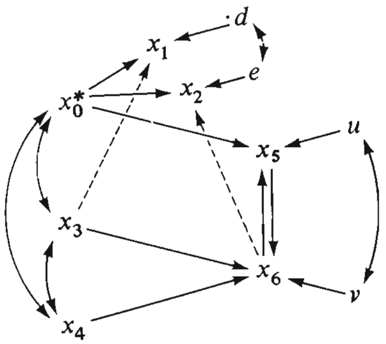
Despite these gains, the contribution of multiple indicators should not be exaggerated. Their help in connection with the identification problem may be one-sided. We have seen that each of the indicators (x_1 and x_2) of the unobserved variable is eligible as an instrument in estimating the equation in which the unobserved variable x_0^* does *not* appear, that is, the x_6-equation. Thus, if the x_6-equation had been underidentified when there was only one indicator of x_0^*, obtaining additional, "well-behaved," indicators would enable us to identify that equation. Note that this equation becomes identified when there are enough indicators, regardless of whether there are correlations among the errors in the indicators. (Recall that our use of x_1 and x_2 as instrumental variables for the x_6-equation did not depend on whether $\sigma_{de} = 0$.) However, when we come to an equation in which the unobserved variable appears as an explanatory variable (a direct cause of the endogenous variable), the situation is different. As we saw, our list of instruments for the x_5-equation depends on whether or not $\sigma_{de} = 0$. More generally, if the errors in the indicators are uncorrelated, we gain instrumental variables by adding (well-behaved) indicators. But if they are all intercorrelated, there is no such gain. Having more indicators will not, in this case, lead to the identification of a hitherto underidentified equation.

The assumption that errors in indicators are uncorrelated may, therefore, be quite useful. But it is also a very strong assumption. With only two indicators there is no way to test it. When there are three (all "well-behaved"), the only possibility of rejecting the assumption of uncorrelated errors is that the correlations among the three indicators will fall into the pattern of the "Heywood case" (Harman, 1967, pages 117–118). With four or more indicators, the hypothesis that all the indicators are "well-behaved" (in that they have no common causes other than the single unobserved variable x_0^*) and that their errors are uncorrelated is tantamount to the hypothesis that their intercorrelations are due to a single common factor. (See Harman, 1967, on factor analysis in general and Duncan, 1972, for a few simple cases.) As a matter of experience sociologists seldom have large sets of indicators with these attractive properties. As an empirical rule (exceptions granted), therefore, it is unlikely that an underidentified equation wherein an unobserved variable is one of the explicit causes will be rendered identifiable by finding several indicators of that variable.

Issues of identification aside, we have also seen that use of multiple indicators gives rise to complications in estimation. These may seem relatively benign, however, if one is able to take advantage of the new methods of estimation specifically contrived to deal with them (Jöreskog, 1970; Werts, Jöreskog, and Linn, 1973). But introduction of the additional indicator(s) does use resources that might be put into alternative use (for example, increasing sample size, or strengthening the model on the substantive side). In this connection, it should be remembered that with only a single indicator, even a highly fallible one (a high value of $\hat{\sigma}_{dd}/\hat{\sigma}_{11}$), we still do not have to suppose that its fallibility is imparting a bias to our estimates of structural coefficients. Only the precision, not the "validity," of those estimates is improved by lengthening our roster of indicators—provided, of course, the one indicator is "well behaved."

So much turns on the question of whether an indicator, or rather, its error structure, is "well behaved" that we should explore more systematically what this means. In essence, we are looking at the problem of "specification error" again, but this time with special reference to the problem of unobserved variables.

Let us complicate the illustrative model we have been working with so as to exemplify the case of indicators that are not "well behaved":



We will consider models including either the coefficient b_{13} or b_{26} or both, as will be made explicit by the x_1- and x_2-equations. The x_5 and x_6 equations remain as they were.

We begin with

$$x_1 = x_0^* + b_{13}x_3 + d$$

$$x_2 = ax_0^* + e$$

(blanking out the arrow from x_6 to x_2). Now, x_1 is no longer a "clean" (if fallible) measure of x_0^*. Instead it is "contaminated" by another exogenous variable in the model. We see why "validity of indicators" must be investigated with reference to a particular model. If x_3 were not in the model or were uncorrelated with x_0^*, this form of contamination would not matter a great deal. We find, perhaps surprisingly, that it is not necessary to alter the previously stated specifications on the errors and disturbances. Indeed, most of our results carry over from the previous version of the model. The x_6-equation is overidentified, with $x_1, \ldots, x_4$ eligible as instruments. The x_5-equation is also overidentified, in the somewhat different sense that there is more than one way to "solve out" x_0^* in that equation, although the two exogenous variables, x_3 and x_4, are still available as instruments. Upon solving out x_0^* from the x_1- and x_2-equations, we have

$$x_2 = ax_1 - ab_{13}x_3 + e'$$

where

$$e' = e - ad$$

or

$$x_1 = \frac{1}{a}x_2 + b_{13}x_3 + d'$$

where

$$d' = d - \frac{e}{a}$$

and thus encounter the indeterminacy in two-stage estimation again, owing to the sensitivity of this method to the normalization rule. As before, optimal methods of estimation are beyond our scope, although it is clear that four instrumental variables ($x_3, \ldots, x_6$) are available. The overidentifying restrictions are made more explicit by the following derivations from the x_1- and x_2-equations:

$$\sigma_{13} = \sigma_{03}^* + b_{13}\sigma_{33} \qquad \sigma_{23} = a\sigma_{03}^*$$

$$\sigma_{14} = \sigma_{04}^* + b_{13}\sigma_{34} \qquad \sigma_{24} = a\sigma_{04}^*$$

$$\sigma_{15} = \sigma_{05}^* + b_{13}\sigma_{35} \qquad \sigma_{25} = a\sigma_{05}^*$$

$$\sigma_{16} = \sigma_{06}^* + b_{13}\sigma_{36} \qquad \sigma_{26} = a\sigma_{06}^*$$

The simple proportionality of the earlier model no longer holds. Instead, we have

$$\frac{\sigma_{13}}{\sigma_{23}} = \frac{1}{a} + \frac{b_{13}\sigma_{33}}{\sigma_{23}}$$

$$\frac{\sigma_{14}}{\sigma_{24}} = \frac{1}{a} + \frac{b_{13}\sigma_{34}}{\sigma_{24}}$$

$$\frac{\sigma_{15}}{\sigma_{25}} = \frac{1}{a} + \frac{b_{13}\sigma_{35}}{\sigma_{25}}$$

$$\frac{\sigma_{16}}{\sigma_{26}} = \frac{1}{a} + \frac{b_{13}\sigma_{36}}{\sigma_{26}}$$

putting the overidentifying restrictions in the form of an (exact) linear relation of the left-hand ratio of covariances to the right-hand ratio. This makes informal inspection of sample evidence bearing on the validity of the overidentifying restrictions easy to carry out. But a formal test is beyond our scope.

What one might find, therefore, is that this model offers an improved specification of the mechanism giving rise to the indicators—supposing, that is, that the first model was found wanting on that score while the present one is acceptable. The estimates for the x_5-equation would presumably be improved as a consequence. But two-stage estimates of the x_6-equation would not be affected. (Other methods—the so-called "system" methods, as distinct from a "single-equation" method like 2SLS or the two-stage method mentioned in this chapter—would, however, be favorably affected by the corrected specification of the model.)

We consider next the version of our illustrative model in which the contamination of an indicator arises from an endogenous rather than an exogenous variable (ignore the dashed arrow from x_3 to x_1 while retaining the one from x_6 to x_2):

$$x_1 = x_0^* + d$$

$$x_2 = ax_0^* + b_{26}x_6 + e$$

"Solving out" x_0^*,

$$x_2 = ax_1 + b_{26}x_6 + e'' \quad \text{where} \quad e'' = e - ad$$

We still have four variables ($x_3, \ldots, x_6$) eligible as instruments. But for the reason noted before, a more complex method of estimation than two-stage regression will be desirable. Overidentifying restrictions are implied by

$$\sigma_{13} = \sigma_{03}^* \qquad \sigma_{23} = a\sigma_{03}^* + b_{26}\sigma_{36}$$

$$\sigma_{14} = \sigma_{04}^* \qquad \sigma_{24} = a\sigma_{04}^* + b_{26}\sigma_{46}$$

$$\sigma_{15} = \sigma_{05}^* \qquad \sigma_{25} = a\sigma_{05}^* + b_{26}\sigma_{56}$$

$$\sigma_{16} = \sigma_{06}^* \qquad \sigma_{26} = a\sigma_{06}^* + b_{26}\sigma_{66}$$

Again, we see that there must be an exact linear relationship between certain ratios of covariances:

$$\frac{\sigma_{23}}{\sigma_{13}} = a + \frac{b_{26}\sigma_{36}}{\sigma_{13}}$$

$$\frac{\sigma_{24}}{\sigma_{14}} = a + \frac{b_{26}\sigma_{46}}{\sigma_{14}}$$

$$\frac{\sigma_{25}}{\sigma_{15}} = a + \frac{b_{26}\sigma_{56}}{\sigma_{15}}$$

$$\frac{\sigma_{26}}{\sigma_{16}} = a + \frac{b_{26}\sigma_{66}}{\sigma_{16}}$$

Informal testing of the overidentifying restrictions is straightforward, but a formal test is beyond our scope.

The changed pattern of contamination has not affected the identification of the x_5-equation; x_3 and x_4 remain eligible as instruments. But we shall still have to resolve the implicit issue of weighting encountered when "solving out" x_0^* from the x_5-equation.

In regard to the x_6-equation, we no longer have $E(x_2 v) = 0$. Instead, as we find by multiplying through the new x_2-equation by v,

$$E(x_2 v) = aE(x_0^* v) + b_{26}E(x_6 v) + E(ev)$$

$$= b_{26}\sigma_{6v}$$

(since the other two covariances are still zero). Therefore, x_2 is not available as an instrument. But this is not fatal, since the x_6-equation was originally overidentified. Loss of one instrumental variable makes it just identified. We may estimate its coefficients by the IV method, with x_1, x_3, and x_4 as instruments.

We have seen that distinct consequences arise, according to whether the source of contamination is an exogenous or an endogenous variable. The strategy for exploring these consequences should now be familiar enough for you to attack the following

Exercise. *Consider a model represented by the previous diagram, but now suppose that both b_{13} and b_{26} are present. Investigate identifiability of the x_5- and x_6-equations and the parameters that govern the behavior of the indicators x_1 and x_2. Point out features of the model that give rise to problems in estimation that cannot be solved by the two-stage procedure suggested earlier and thus go beyond the scope of this book.*

The examples and exercise cover only a few special cases. The reader should educate his intuition concerning problems of this kind by designing models to exemplify other cases. For example, how are the identification and estimation of the x_6-equation affected if x_5 is the source of contamination of one of the indicators? It seems possible to generalize to this extent. If a model is heavily overidentified, then quite a bit of quite "nasty" contamination of exogenous variables (only exogenous variables have been considered so far) is tolerable. However, every instance of contamination exacts a price in terms of the number of instrumental variables and number of overidentifying restrictions. The consequences, however, are not necessarily the same for the several equations of the model. Both the effect on identification and the implications for a strategy of estimation have to be studied carefully for each new model. Another caution: A situation in which the two (or more) indicators have the same pattern of contamination is tricky. Suppose, for example, that both x_1 and x_2 are functions of x_3 as well as x_0^*; our equations would read:

$$x_1 = x_0^* + b_{13}x_3 + d$$

$$x_2 = ax_0^* + b_{23}x_3 + e$$

"Solving out" x_0^* we have

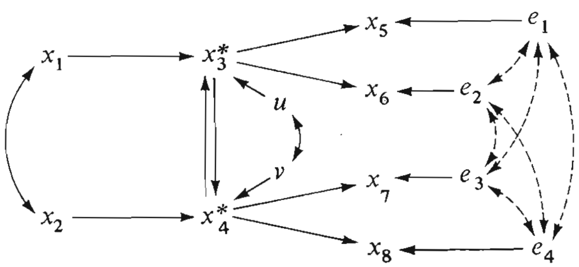
$$x_2 = ax_1 - ab_{13}x_3 + b_{23}x_3 + e - ad$$

$$= ax_1 + (b_{23} - ab_{13})x_3 + e - ad$$

We can hope to estimate only one distinct coefficient for x_3 (even with the more advanced methods that are beyond the scope of this book!). Hence, we cannot distinguish between b_{23} and ab_{13}.

Exercise. *Determine whether this gives rise to unmanageable problems in connection with estimating the x_5- and x_6-equations.*

To illustrate complications that arise from multiple indicators of endogenous variables, we modify the model in Chapter 5 in such a way that each endogenous variable is unobserved but has two indicators:



Our theory of how the indicators behave is expressed in four equations

$$x_5 = x_3^* + e_1$$

$$x_6 = a_1 x_3^* + e_2$$

$$x_7 = x_4^* + e_3$$

$$x_8 = a_2 x_4^* + e_4$$

We leave open temporarily the question of whether the errors should be assumed to be uncorrelated. The substance of the model is given by the two nonrecursive equations:

$$x_3^* = b_{31} x_1 + b_{34} x_4^* + u$$

$$x_4^* = b_{42} x_2 + b_{43} x_3^* + v$$

Exogenous variables are specified to have zero covariances with disturbances and measurement errors. Also, the unobserved variables have zero covariances with measurement errors. Hence (verify this) $E(e_h u) = E(e_h v) = 0$, $h = 1, \ldots, 4$; but $E(uv) \neq 0$.

We can easily "solve out" x_3^* from the x_5- and x_6-equations (the parallel discussion for x_4^* in the x_7- and x_8-equations can be supplied by relabelling subscripts):

$$x_6 = a_1 x_5 + e_2 - a_1 e_1$$

Two instrumental variables are available (x_1 and x_2), so that a_1 is overidentified. But in view of the indeterminacy of the normalization,

two-stage regression has drawbacks as a method of estimation. Making explicit the overidentifying restrictions:

$$\sigma_{15} = \sigma_{13}^* \qquad \sigma_{16} = a_1 \sigma_{13}^*$$

$$\sigma_{25} = \sigma_{23}^* \qquad \sigma_{26} = a_1 \sigma_{23}^*$$

$$\sigma_{17} = \sigma_{14}^* \qquad \sigma_{18} = a_2 \sigma_{14}^*$$

$$\sigma_{27} = \sigma_{24}^* \qquad \sigma_{28} = a_2 \sigma_{24}^*$$

Hence,

$$a_1 = \frac{\sigma_{16}}{\sigma_{15}} = \frac{\sigma_{26}}{\sigma_{25}}$$

$$a_2 = \frac{\sigma_{18}}{\sigma_{17}} = \frac{\sigma_{28}}{\sigma_{27}}$$

A simple informal test is obvious, but a formal test is beyond our scope.

At this point, let us consider the implications of the specification of uncorrelated measurement errors, $E(e_h e_j) = 0$; $h, j = 1, \ldots, 4$; $h \neq j$. We multiply each of the four indicator equations by each of the others, to obtain:

$$\sigma_{56} = a_1 \sigma_{33}^* + \sigma_{e_1 e_2}$$

$$\sigma_{78} = a_2 \sigma_{44}^* + \sigma_{e_3 e_4}$$

$$\sigma_{57} = \sigma_{34}^* + \sigma_{e_1 e_3}$$

$$\sigma_{58} = a_2 \sigma_{34}^* + \sigma_{e_1 e_4}$$

$$\sigma_{67} = a_1 \sigma_{34}^* + \sigma_{e_2 e_3}$$

$$\sigma_{68} = a_1 a_2 \sigma_{34}^* + \sigma_{e_2 e_4}$$

Setting all the error covariances to zero, there are five distinct parameters on the right-hand side and six covariances of observable variances on the left-hand side. The parameters are clearly overidentified. What we have in our four equations is a particular factor analysis model, methods of estimation for which would have to be sought in the literature of that subject. We see, however, that estimates of a_1 and a_2

obtained by this route will not be the same as those obtained by using instrumental variables. That is, still another kind of overidentifying restriction is implied by the availability of these two distinct solutions. At this point, an investigator might well ask herself whether she needs so strong a theory about her indicators. Relinquishing the assumption of uncorrelated measurement errors will cost her the possibility of estimating the variances and covariance of the unobserved variables. But there is no clear need for these estimates in any case, since the structural coefficients in the x_3^*- and x_4^*-equations can be estimated without knowing them. Henceforth, we shall assume that any pair of e's may have a nonzero covariance.

Turning to our structural equations, we see that the estimation problem is complicated by the fact that both x_3^* and x_4^* appear in both equations and that there are alternative ways to "solve out" the unobserved variables. On the other hand, it is clear that once this is done—we will assume that the appropriate "weighting" of the x_5-vis-à-vis the x_6-equation and the x_7- vis-à-vis the x_8-equation will, in effect, be determined in the process of estimating a_1 and a_2—each equation is identified, since there are two structural coefficients in each and two instrumental variables are available. As is the case throughout this chapter, we consider that the reader has fair warning if we state that a method of estimation substantially more complicated than two-stage regression will be required for this problem.

The advantages and disadvantages of multiple indicators for endogenous variables turn out to be much the same as those noted for exogenous variables earlier in the chapter. On the one hand, they do not necessarily help with the identification problem, depending on what one can assume about correlations among errors in the indicators. A model that is underidentified with one indicator per endogenous variable may remain so with many indicators. Estimation is more complicated and costly than with a single indicator. On the other hand, one does secure the chance to make a partial test of whether the indicators are "well behaved." If it turns out that they are, estimates of structural coefficients are more precise (ordinarily they will have smaller sampling errors) than they would be with one indicator per variable. But there is no other sense in which those estimates are improved.

However fascinating may be the problems that arise with multiple indicators, we have to recognize that sometimes they merely complicate if not obscure what is surely the more fundamental problem: proper specification of our models in substantive terms.

FURTHER READING

In this chapter it has been necessary to mention repeatedly "methods that go beyond the scope of this book." The most important references are two papers of Jöreskog (1970, and his contribution to Goldberger and Duncan, 1973, Chapter 5). These are highly technical articles. A brief and more accessible exposition, with several examples, is given by Werts, Jöreskog, and Linn (1973). For a critique of earlier sociological writing on the multiple indicators problem and an explication of the meaning of "efficient estimation" in this context, see Hauser and Goldberger (1971).