ture correlations, in subsequent chapters both recursive and nonrecursive models are treated from a different and more fundamental point of view.

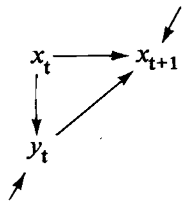
FURTHER READING

The classic exposition of the causal structures that may underlie the correlations among three variables is that of Simon (1954). A didactic presentation of four-variable models much in the spirit of this chapter is given by Blalock (1962–1963). Both papers are reprinted in Blalock (1971).

3

Recursive Models

A model is said to be recursive if all the causal linkages run "one way," that is, if no two variables are reciprocally related in such a way that each affects and depends on the other, and no variable "feeds back" upon itself through any indirect concatenation of causal linkages, however circuitous. However, recursive models do cover the case in which the "same" variable occurs at two distinct points in time, for, in that event, we would regard the two measurements as defining two different variables. For example, a dynamic model like the following:



where t and $t + 1$ are two points in time, is recursive (even though x appears to feed back upon itself). The definition also subsumes the case in which there are two or more ostensibly contemporaneous dependent variables where none of them has a direct or indirect causal linkage to any other. This situation is illustrated by Models II and II′ in Chapter 2 (assuming that ρ_{uv} in Model II′ does not implicitly arise from either a path $y \to z$ or a path $z \to y$). In this case, we have to consider whether or not to specify $\rho_{uv} = 0$. If so, we might term the model "fully recursive"; if not, it is merely "recursive."

With the exception of this last kind of situation—which offers no difficulty in principle, though it requires careful handling in practice—we can state that all the *dependent variables* in a recursive model (those whose causes are explicitly represented in the model) are arrayed in an unambiguous causal ordering. Moreover, all *exogenous variables* (those whose causes are not explicitly represented in the model) are, as a set, causally prior to all the dependent variables. There is, however, no causal ordering of the exogenous variables (if there are two or more) with respect to each other. (In some other model, of course, these variables might be treated as dependent variables.)

It simplifies matters greatly and results in a more powerful model if we can assume there is only one exogenous variable. This may not always be a reasonable assumption. We will consider models of this kind first and then see what modifications are entailed if we have to assume the contrary.

All our exposition of recursive models will rest on illustrations in which there are just four variables. Throughout this book we rely on relatively simple examples, and it is expected that the reader will come to see how the principles pertaining to these examples can be generalized to other models. There is some risk in this procedure, but it is hopefully outweighed by the advantages of making the discussion both concrete and compact—virtues difficult to attain if all principles and theorems must be stated in a perfectly general form.

With four variables, one of them exogenous, the causal ordering will involve a unique arrangement. Any one of the four might be the exogenous variable, any of the remaining three might be the first dependent variable, and either of the remaining two might follow it. Hence, there are $4 \cdot 3 \cdot 2 \cdot 1 = 24$ possible causal orderings of four variables. To build a recursive model means that we must choose one and only one of these 24 as the "true" ordering. That choice, it should be clear from the discussion in the previous chapter, cannot be based on the correlations among the variables, because *any* correlation matrix will be consistent with *any* causal ordering one may propose.

The information in regard to causal ordering is, logically, a priori. Such information is derived from theory, broadly construed, and no amount of study of the formal properties of models can teach one how to come up with a true theory. We can, however, prescribe the task of theory—to provide a causal ordering of the variables. Another way to put it is that the theory must tell us that at least six of the twelve possible causal linkages among four variables are *not* present, and these missing links, moreover, must fall into a triangular pattern. For, if four variables are put into a causal order and then numbered in sequence, we will have a pattern like the following:

	Caused by			
Effect	x_1	x_2	x_3	x_4
(x_1)	...	0	0	0
x_2	×	...	0	0
x_3	×	×	...	0
x_4	×	×	×	...

It is immaterial whether one or more of the crosses is replaced by a nought. But all the noughts must be present, to stand for the assumption that x_2 does *not* cause x_1 (though x_1 may cause x_2), and so on. If one enters six noughts in such a matrix (ignoring diagonal cells) before the variables are numbered, it may or may not be possible to triangulate the matrix. If it is, then the matrix defines a recursive causal ordering. If not, one or more nonrecursive relationships is present. Thus, the indispensable contribution of theory is to put noughts into the matrix. It is an odd way to put it, but the decisive criterion of the utility of a theory is that it can tell us definitely what causal relationships do not obtain, not that it can suggest (however evocatively) what relationships may well be present. (This remark will seem even more poignant when, in a later chapter, we discuss the problem of identification for a nonrecursive model.)

There is still another way to describe a recursive model. We may say that all exogenous variables in the model are *predetermined* with respect to all dependent variables. Moreover, each dependent variable is predetermined with respect to any other dependent variable that occurs later in a causal ordering. We take this to mean that all exogenous variables are uncorrelated with the disturbances in all equations of the model. (If one cannot assume this, the remedy is to build a better model.) Moreover, we shall similarly assume that the predetermined variables occurring in any equation are uncorrelated with the disturbance of that equation. (This is virtually a definition of "predetermined." It does leave open the possibility that of two dependent variables in a recursive model, neither is predetermined with respect to the other, while the disturbances of their equations are correlated, as is the case for Model II′ in the previous chapter.) Again, this assumption has to be evaluated on its theoretical or substantive merits and, if it must be faulted, the recourse is to propose a better model.

The four-variable model, already described by a matrix of crosses and noughts, is more explicitly represented by this set of equations:

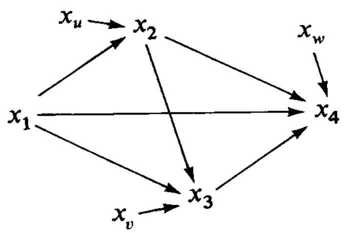
(x_1 exogenous)

$$x_2 = p_{21}x_1 + p_{2u}x_u$$

$$x_3 = p_{32}x_2 + p_{31}x_1 + p_{3v}x_v$$

$$x_4 = p_{43}x_3 + p_{42}x_2 + p_{41}x_1 + p_{4w}x_w$$

or by the path diagram:



We continue to assume that $E(x_h) = 0$ and $E(x_h^2) = 1$, $h = 1,2,3,4,$ u,v,w (the variables are in standard form). Hence, $E(x_h x_j) = \rho_{hj}$, the correlation (in the population) between x_h and x_j. The specifications on the disturbance terms are as follows:

(*a*) The exogenous variable is uncorrelated with the disturbances:

$$E(x_1 x_u) = E(x_1 x_v) = E(x_1 x_w) = 0$$

(*b*) Disturbances are also uncorrelated with any other predetermined variables in an equation:

$$E(x_2 x_v) = E(x_2 x_w) = E(x_3 x_w) = 0$$

A standard manipulation is to "multiply through" one equation of the model by a variable in the model, take expected values, and express in terms of path coefficients (the p's) and correlations (ρ's). Following this procedure with the x_2-equation, making use of (*a*) and (*b*), we deduce that the disturbance in each equation is uncorrelated with the disturbance in any other equation; for example,

$$E(x_2 x_v) = p_{21}E(x_1 x_v) + p_{2u}E(x_u x_v)$$

so that $\rho_{uv} = 0$; similarly, $\rho_{uw} = \rho_{vw} = 0$. But note that the disturbance in each equation has a nonzero correlation with the dependent variable in that equation and (in general) with the dependent variable in each "later" equation. To take another example, if we multiply through the equation for x_3 by x_2, we obtain

$$E(x_2 x_3) = p_{32}E(x_2^2) + p_{31}E(x_1 x_2) + p_{3v}E(x_2 x_v)$$

or

$$\rho_{23} = p_{32} + p_{31}\rho_{12}$$

[since $E(x_2^2) = 1$ and $E(x_2 x_v) = 0$]. Proceeding systematically, we obtain

$$\rho_{12} = p_{21} \qquad \text{(from the } x_2\text{-equation)}$$

$$\left.\begin{aligned} \rho_{13} &= p_{31} + p_{32}\rho_{12} \\ \rho_{23} &= p_{31}\rho_{12} + p_{32} \end{aligned}\right\} \qquad \text{(from the } x_3\text{-equation)}$$

$$\left.\begin{aligned} \rho_{14} &= p_{41} + p_{42}\rho_{12} + p_{43}\rho_{13} \\ \rho_{24} &= p_{41}\rho_{12} + p_{42} + p_{43}\rho_{23} \\ \rho_{34} &= p_{41}\rho_{13} + p_{42}\rho_{23} + p_{43} \end{aligned}\right\} \qquad \text{(from the } x_4\text{-equation)}$$

We may study this set of "normal equations" in two ways:

(*i*) *Solve for the p's in terms of the ρ's*

We obtain, for the first equation of the model,

$$p_{21} = \rho_{12}$$

for the second equation,

$$p_{31} = (\rho_{13} - \rho_{12}\rho_{23})/(1 - \rho_{12}^2)$$
$$p_{32} = (\rho_{23} - \rho_{12}\rho_{13})/(1 - \rho_{12}^2)$$

and for the third equation,

$$p_{41} = \frac{1}{D}\begin{vmatrix} \rho_{14} & \rho_{12} & \rho_{13} \\ \rho_{24} & 1 & \rho_{23} \\ \rho_{34} & \rho_{23} & 1 \end{vmatrix}$$

$$p_{42} = \frac{1}{D}\begin{vmatrix} 1 & \rho_{14} & \rho_{13} \\ \rho_{12} & \rho_{24} & \rho_{23} \\ \rho_{13} & \rho_{34} & 1 \end{vmatrix}$$

$$p_{43} = \frac{1}{D}\begin{vmatrix} 1 & \rho_{12} & \rho_{14} \\ \rho_{12} & 1 & \rho_{24} \\ \rho_{13} & \rho_{23} & \rho_{34} \end{vmatrix}$$

where

$$D = \begin{vmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{vmatrix}$$

Thus, if we knew the correlations we could solve for the coefficients of the model. In practice, we have only estimates (from a sample) of the correlations. If these estimates are inserted into the foregoing formulas (in place of the population correlations), the formulas will yield estimates of the p's. These are, in fact, the same estimates that one obtains from the ordinary least squares (OLS) regression of

x_2 on x_1

x_3 on x_2 and x_1

x_4 on x_3, x_2, and x_1

if the variables are in standard form.

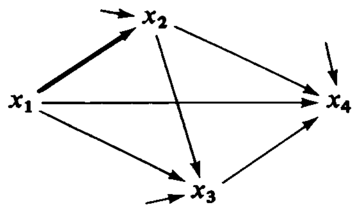
(ii) Solve for the ρ's in terms of the p's

This may be done quite simply, making substitutions in the "normal equations." We obtain

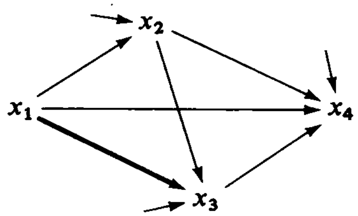
$$\rho_{12} = p_{21}$$
$$\rho_{13} = p_{31} + p_{32}p_{21}$$
$$\rho_{23} = p_{32} + p_{31}p_{21}$$
$$\rho_{14} = p_{41} + p_{42}p_{21} + p_{43}(p_{31} + p_{32}p_{21})$$
$$\rho_{24} = p_{42} + p_{43}p_{32} + p_{41}p_{21} + p_{43}p_{31}p_{21}$$
$$\rho_{34} = p_{43} + p_{42}(p_{32} + p_{31}p_{21}) + p_{41}(p_{31} + p_{32}p_{21})$$

These expressions are instructive in that they show how the model generates the (observable) correlations. It is worthwhile to study them with some care.

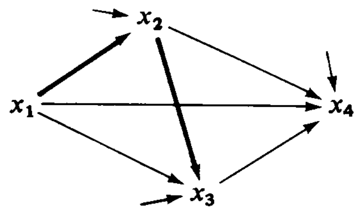
(*1*) We see that the entirety of the correlation between x_1 and x_2 is generated by the direct effect, p_{21}



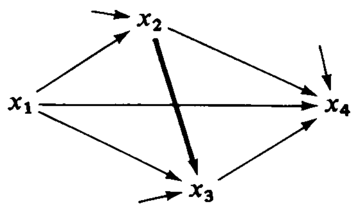
(*2*) The correlation between x_1 and x_3 is generated by two distinct paths, so that ρ_{13} equals the direct effect, p_{31}



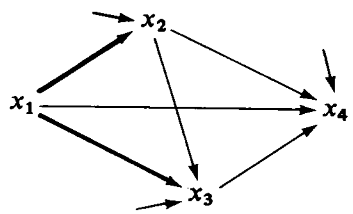
plus the indirect effect, $p_{32}p_{21}$



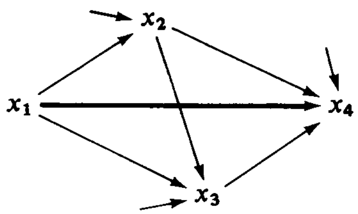
(*3*) The situation is different in regard to x_2 and x_3, for here we have the total correlation (ρ_{23}) generated as the sum of the direct effect, p_{32}



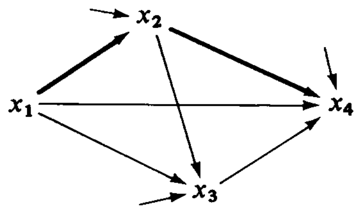
plus correlation due to a common cause, $p_{31}p_{21}$



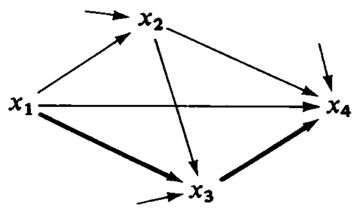
(*4*) The correlation between x_1 and x_4 is generated by four distinct causal links; ρ_{14} equals the direct effect, p_{41}



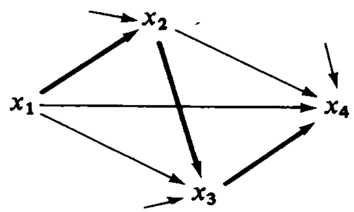
plus indirect effect via x_2, $p_{42}p_{21}$



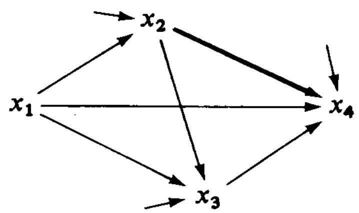
plus indirect effect via x_3, $p_{43}p_{31}$



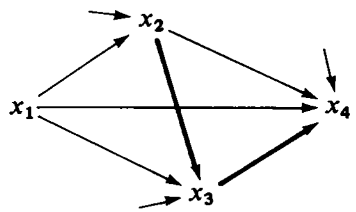
plus indirect effect via x_3 and x_2, $p_{43}p_{32}p_{21}$



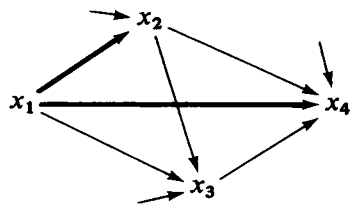
(*5*) Both an indirect effect and correlation due to common causes are involved in generating the correlation between x_2 and x_4; ρ_{24} equals the direct effect, p_{42}



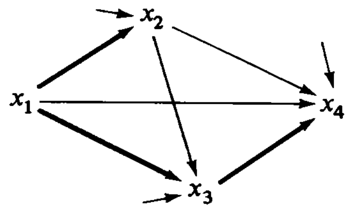
plus indirect effect via x_3, $p_{43}p_{32}$



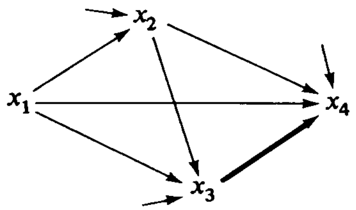
plus correlation due to x_1 operating as a common cause, directly, $p_{41}p_{21}$



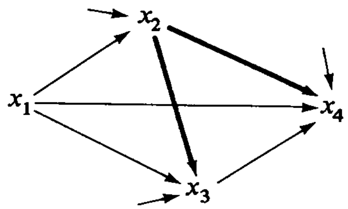
and indirectly (via x_3), $p_{43}p_{31}p_{21}$



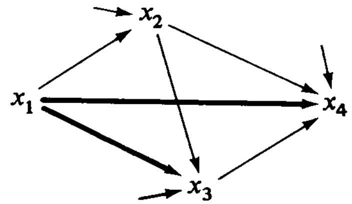
(*6*) There are no indirect effects in the model producing correlation between x_3 and x_4; but there are two common causes. Hence, ρ_{34} equals the direct effect, p_{43}



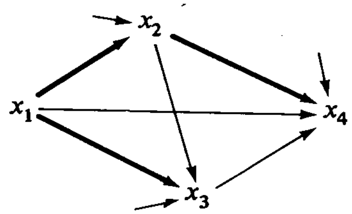
plus correlation due to common causes, working directly, $p_{42}p_{32}$



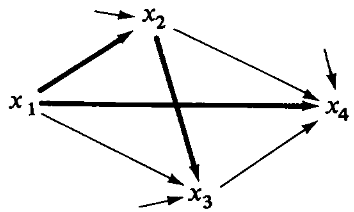
and $p_{41}p_{31}$



or indirectly, $p_{42}p_{31}p_{21}$



and $p_{41}p_{32}p_{21}$



In all six of these cases, the correlation may be read off the path diagram using Wright's (1921) multiplication rule: To find the correlation between x_h and x_j, where x_j appears "later" in the model, begin at x_j and read *back* to x_h along each distinct direct and indirect (compound) path, forming the product of the coefficients along that path. After reading back, read *forward* (if necessary), but only one reversal from back to forward is permitted. Sum the products obtained for all the linkages between x_j and x_h.

We consider next a modification of the model. Suppose both x_1 and x_2 are exogenous, that is, the model cannot explain how they are generated. Since we do not know anything about this, we cannot make any strong assumption about their correlation. Hence, we shall have to suppose, in general, that $\rho_{12} \neq 0$. The model now has but two equations,

$$x_3 = p_{32}x_2 + p_{31}x_1 + p_{3v}x_v$$

$$x_4 = p_{43}x_3 + p_{42}x_2 + p_{41}x_1 + p_{4w}x_w$$

Disturbances are uncorrelated with predetermined (including exogenous) variables. Hence, $\rho_{1v} = \rho_{1w} = \rho_{2v} = \rho_{2w} = \rho_{3w} = 0$; and, as a consequence, $\rho_{vw} = 0$. Normal equations are obtained as before, except, of course, that there is no normal equation with ρ_{12} on the left-hand side. Also, except for the fact that no p_{21} occurs in the model, the solution for the p's is the same as before and has the same interpretation.

In studying the model from standpoint *ii* (solution for ρ's in terms of p's), however, we must reconsider the situation. Algebraic substitutions yield,

$$\rho_{13} = p_{31} + p_{32}\rho_{12}$$

$$\rho_{23} = p_{32} + p_{31}\rho_{12}$$

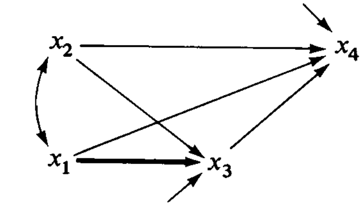
$$\rho_{14} = p_{41} + p_{43}p_{31} + (p_{42} + p_{43}p_{32})\rho_{12}$$

$$\rho_{24} = p_{42} + p_{43}p_{32} + (p_{41} + p_{43}p_{31})\rho_{12}$$

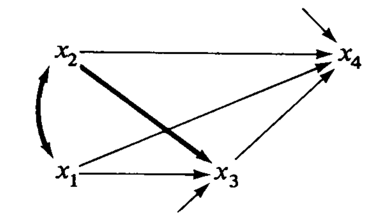
$$\rho_{34} = p_{43} + p_{42}p_{32} + p_{41}p_{31} + (p_{42}p_{31} + p_{41}p_{32})\rho_{12}$$

We cannot eliminate ρ_{12} from the right-hand side of these equations, since the model cannot (by definition) tell us anything about how that

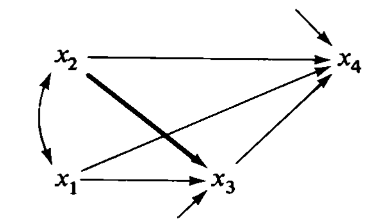
correlation is generated. Moreover, the presence of this correlation means that the remaining correlations are generated in a somewhat ambiguous way. Consider the correlation between x_1 and x_3. We have a direct effect, p_{31}. The other term, $p_{32}\rho_{12}$, consists of the product of the direct effect of x_2 on x_3 and the correlation of x_1 and x_2. It represents a contribution to ρ_{13} by virtue of the fact that *another cause* of x_3 (namely x_2) is correlated (to the extent of ρ_{12}) with the cause we are examining at the moment (namely, x_1). As in Chapter 2 (see Models II′ and III) we use a curved, double-headed arrow to refer to a correlation that cannot be analyzed in terms of causal components within this model. Hence the way to look at this situation is that ρ_{13} equals the direct effect, p_{31}



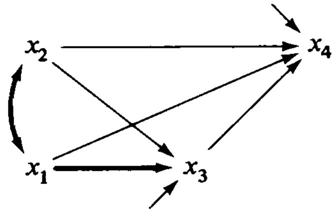
plus correlation due to correlation with another cause, $p_{32}\rho_{12}$



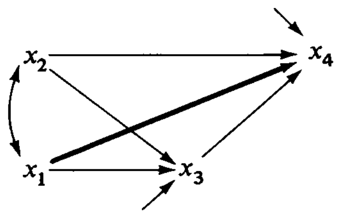
Similarly, we may break down ρ_{23} into the components: direct effect, p_{32}



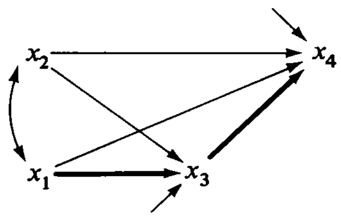
plus correlation due to correlation with another cause, $p_{31}\rho_{12}$



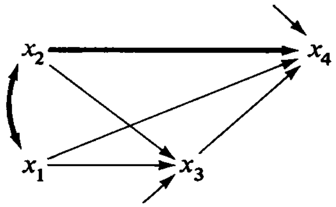
The same kind of reasoning interprets the remaining correlations. For ρ_{14} we obtain the direct effect, p_{41}



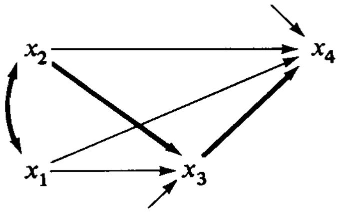
plus the indirect effect, $p_{43}p_{31}$



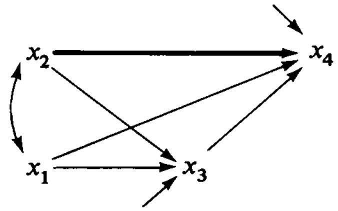
plus correlation due to the correlation of x_1 with another cause (x_2), working both directly, $p_{42}\rho_{12}$



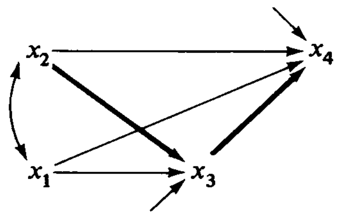
and indirectly, $p_{43}p_{32}\rho_{12}$



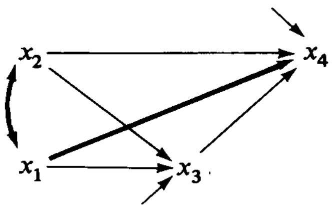
We decompose ρ_{24} into the direct effect, p_{42}



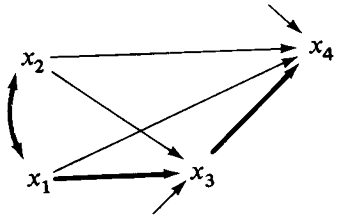
plus the indirect effect, $p_{43}p_{32}$



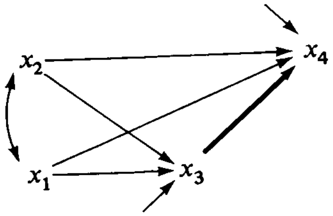
plus correlation due to the correlation of x_2 with another cause (x_1), working both directly, $p_{41}\rho_{12}$



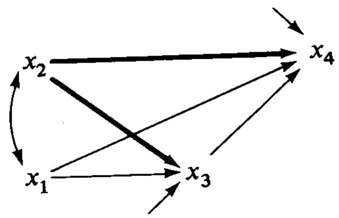
and indirectly, $p_{43}p_{31}\rho_{12}$



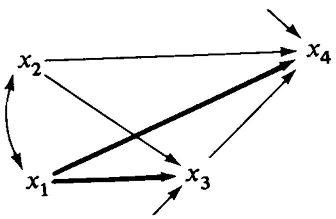
As in the previous version of the model, ρ_{34} involves no indirect effects, but the correlation generated by the common causes (x_1 and x_2) involves both the direct effects of those causes and the correlation due to the fact that they are correlated with each other. Hence, ρ_{34} equals the direct effect, p_{43}



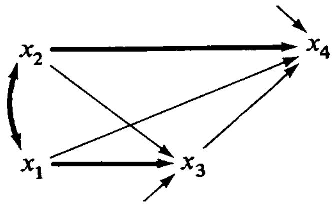
plus correlation due to x_2 as a common cause, $p_{42}p_{32}$



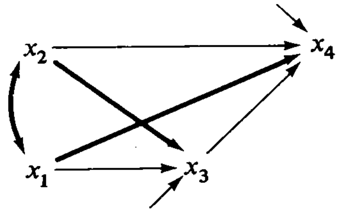
and x_1 as a common cause, $p_{41}p_{31}$



plus correlation due to the correlation of x_1 with another common cause (x_2), $p_{42}p_{31}\rho_{12}$



and correlation due to the correlation of x_2 with another common cause (x_1), $p_{41}p_{32}\rho_{12}$



For some purposes, one might be content to aggregate the last four components, so as to describe the correlation ρ_{34} as being generated by the direct effect (p_{43}) and the correlation due to the influence of the two common causes, x_1 and x_2, on x_3 and x_4.

To bring this discussion within the scope of Sewall Wright's multiplication rule, we stipulate that the curved double-headed arrow is read either back or forward.

It is, of course, an undesirable property of this modified model that we cannot clearly disentangle the effects of its two exogenous variables. But our theory may simply be unable to tell us whether x_1 causes x_2, x_2 causes x_1, each influences the other, both are effects of one or more common or correlated causes, or some combination of these situations holds true. In that event, we cannot know for sure whether a change initiated in (say) x_1 will have indirect effects via x_2 or not, since we do not know whether x_2 depends on x_1. It follows that we cannot, with this model, estimate the *total* effect (defined as *direct* effect plus *indirect* effect) of x_1 on, say, x_4. There may or may not be an indirect causal linkage from x_1 through x_2 to x_4. But since we know nothing about this, we cannot include any such indirect effect in our estimate of total effect.

It should be noted, in both forms of the model, that the zero-order correlation between two variables often is not the correct measure of total effect of one variable on the other, since that correlation may include components other than direct and indirect effects.

In a further modification of the four-variable model, we suppose that x_1, x_2, and x_3 all are exogenous. This gives rise to the degenerate case of a single-equation model:

$$x_4 = p_{43}x_3 + p_{42}x_2 + p_{41}x_1 + p_{4w}x_w$$

The disturbance is uncorrelated with the exogenous variables.

There are three normal equations:

$$\rho_{14} = p_{41} + p_{42}\rho_{12} + p_{43}\rho_{13}$$

$$\rho_{24} = p_{41}\rho_{12} + p_{42} + p_{43}\rho_{23}$$

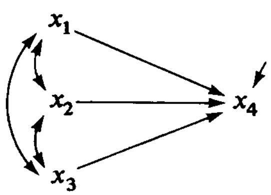
$$\rho_{34} = p_{41}\rho_{13} + p_{42}\rho_{23} + p_{43}$$

As before, if sample correlations are inserted into these equations, the solution for the p's yields least-squares estimates of the parameters.

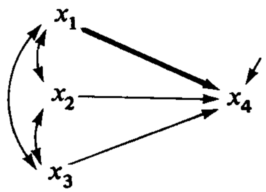
None of the correlations on the right-hand side of the normal equations can be expressed in terms of path coefficients. Therefore, we cannot separate indirect effects from correlation due to common or correlated causes. Thus, the only decomposition we can provide is the following:

Causal variable	*Total correlation*	=	*Direct effect on* x_4	+	*Correlation due to common and/or correlated causes*
x_1	ρ_{14}		p_{41}		$p_{42}\rho_{12} + p_{43}\rho_{13}$
x_2	ρ_{24}		p_{42}		$p_{41}\rho_{12} + p_{43}\rho_{23}$
x_3	ρ_{34}		p_{43}		$p_{41}\rho_{13} + p_{42}\rho_{23}$

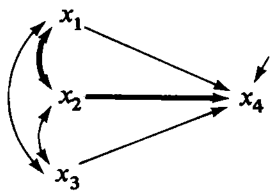
The path diagram for this single-equation model is shown below:



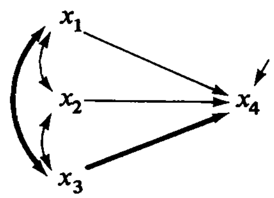
Correlations between exogenous variables are represented by curved, double-headed arrows. The normal equations can be written using Sewall Wright's rule. The curved arrow can be read either forward or backward, but *only one* curved arrow can be included in a given trajectory. Thus, to find ρ_{14} (for example), we read p_{41}



plus $p_{42}\rho_{12}$



plus $p_{43}\rho_{13}$



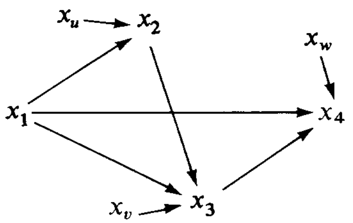
The single-equation model correctly represents the *direct effects* of the exogenous variables. But, since the causal structure of relationships among the exogenous variables is unknown, this model cannot tell us anything about how the indirect effects (not to mention total effects) are generated. In this respect, it is even less satisfactory than the two-equation model.

A major goal of theory, therefore, should be to supply a model that will make some of the exogenous variables endogenous.

This discussion has illustrated a significant theorem: The *direct* effects of predetermined variables in one equation of a model on the dependent variable of that equation are the same, irrespective of the causal relationships holding among the predetermined variables. Thus, returning to the complete model discussed at the beginning, the values of p_{41}, p_{42}, and p_{43} in the third equation do not depend on whether x_2 causes x_1 or vice versa in the first equation or on whether x_1, x_2, or x_3 is the dependent variable in the second equation.

Thus far we have only considered models in which all direct paths allowed by the causal ordering are, in fact, present in the model. We must now consider procedures suited to recursive models in which one or more such paths are (or may be) missing. These procedures have to do with the distinct (though related) problems of *estimation* and *testing*. We discuss them in that order.

We have seen that in the fully recursive model with direct paths from each "earlier" variable to each "later" variable, the path coefficients may be estimated by OLS regression. Suppose, however, that our model, while recursive, explicitly specifies that one or more coefficients are zero. To take a concrete example, consider this path diagram



The equations of the model are

$$(x_1 \text{ exogenous})$$

$$x_2 = p_{21}x_1 + p_{2u}x_u$$

$$x_3 = p_{32}x_2 + p_{31}x_1 + p_{3v}x_v$$

$$x_4 = p_{43}x_3 + p_{41}x_1 + p_{4w}x_w$$

The only change from the model on page 28 is that $p_{42} = 0$. We continue to assume that all variables (including disturbances) are in standard form. In each equation of the model, the disturbance is uncorrelated with the predetermined variables. Moreover, we take the model to be fully recursive, so that the disturbance in each equation is uncorrelated with predetermined variables in all "earlier" equations. (Here, as elsewhere in this chapter, we allow the zero correlation of two disturbances to appear as a consequence of the zero correlation of disturbances with prior predetermined variables.) The force of this specification, with special reference to the present example, is that $\rho_{2w} = 0$, even though x_2 does not appear (explicitly) in the x_4-equation. We have, then, the following specification on the disturbances: $\rho_{1u} = \rho_{1v} = \rho_{1w} = \rho_{2v} = \rho_{2w} = \rho_{3w} = 0$. As a consequence of this specification, we find that it is also true that $\rho_{uv} = \rho_{uw} = \rho_{vw} = 0$.

The normal equations for the x_2-equation and the x_3-equation are the same as before, and OLS estimates of their path coefficients are obtained by formulas given earlier.

Multiplying through the x_4-equation by each predetermined variable, we find

$$\rho_{14} = p_{41} + p_{43}\rho_{13}$$

$$\rho_{24} = p_{41}\rho_{12} + p_{43}\rho_{23}$$

$$\rho_{34} = p_{41}\rho_{13} + p_{43}$$

Assuming the ρ's are known, we have three equations in the two unknown path coefficients. In mathematical terms, the solution for the p's is overdetermined. In the language of structural equation models, the x_4-equation is *overidentified.* In the event that an equation in the model is overidentified, we may deduce that one or more *overidentifying restrictions* must hold if the model is true. Here, we can ascertain the overidentifying restriction by writing out each of the solutions for the p's obtained upon solving a pair of the normal equations. There are three distinct solutions. If the model holds, the values obtained in all three must be equal. Thus,

$$\begin{array}{cccc} & (i) & (ii) & (iii) \\ p_{41} = & \dfrac{\rho_{14} - \rho_{13}\rho_{34}}{1 - \rho_{13}^2} = & \dfrac{\rho_{14}\rho_{23} - \rho_{13}\rho_{24}}{\rho_{23} - \rho_{12}\rho_{13}} = & \dfrac{\rho_{24} - \rho_{23}\rho_{34}}{\rho_{12} - \rho_{13}\rho_{23}} \\ p_{43} = & \dfrac{\rho_{34} - \rho_{13}\rho_{14}}{1 - \rho_{13}^2} = & \dfrac{\rho_{24} - \rho_{12}\rho_{14}}{\rho_{23} - \rho_{12}\rho_{13}} = & \dfrac{\rho_{12}\rho_{34} - \rho_{13}\rho_{24}}{\rho_{12} - \rho_{13}\rho_{23}} \end{array}$$

where solution number (*i*) makes use of the first and third normal equations, number (*ii*) is from the first and second normal equations, and number (*iii*) is from the last two normal equations. If $p_{41}^{(i)} = p_{41}^{(ii)}$ it follows that

$$\rho_{24} + \rho_{13}\rho_{14}\rho_{23} + \rho_{12}\rho_{13}\rho_{34} - \rho_{24}\rho_{13}^2 - \rho_{23}\rho_{34} - \rho_{12}\rho_{14} = 0$$

We note that this expression is just the expansion of the determinant given as the numerator of p_{42} on page 30; and that determinant must be zero if $p_{42} = 0$. We reach the same conclusion from any of the other equalities of solutions for p_{41} or p_{43}. These several equalities are not independent. There is actually only one overidentifying restriction on this model.

Now, the overidentifying restriction must hold in any *population* in which the model applies. But if we have only *sample* values of the correlations we cannot expect it to hold exactly, nor can we expect the three solutions for each path coefficient to be exactly equal. In that event, to estimate the path coefficients we must choose one of the solutions, or perhaps some average of them. Since each solution makes use of only two of the normal equations, it would appear that averaging the solutions would be advisable, since we would then be making use of all the sample correlations rather than only some of them. *This intuition, however, is wrong.* It turns out that the preferred estimate is obtained upon inserting sample correlations into solution (*i*). It will be noted that the estimates of p_{41} and p_{43} obtained in this way are just the OLS regression coefficients of x_4 on x_1 and x_3. The general rule, then, is this: In a fully recursive model (where the correlation between each pair of disturbances is zero), estimate the coefficients in each equation by OLS regression of the dependent variable on the predetermined variables included in that equation.

The basis for this rule is a proof that the sampling variance of a $\hat{p}$ estimated by OLS is smaller than the variance of any other unbiased estimate of the same coefficient, even if such an estimate appears to use more information in the sense of combining correlations involving the included variables with correlations involving the excluded predetermined variable(s). Some of the earlier literature on path analysis was in error on this point; it was called to the attention of sociologists by A. S. Goldberger (1970).

We now turn to the problem of *testing*. In the preceding example, we discussed estimation on the assumption that the model and, in particular, the overidentifying restriction on the model are known in advance

to be true. But the investigator may not feel confident of this specification. Indeed, he may be undertaking a study precisely to test that aspect of his theory which says that a particular coefficient should be zero. Much of the literature on causal models in the 1960s—particularly papers discussing or using the so-called "Simon–Blalock technique"—focused on this very question. Sometimes the problem was described as that of "making causal inferences from correlational data," but that ambiguous phrase seems to promise far too much. In the light of our preceding discussion, it would be more accurate to describe the problem as that of *testing the overidentifying restriction(s)* of a model.

We first take note of two plausible and conceptually correct procedures for making such tests on recursive models. But, since these procedures are not convenient from the standpoint of the standard methods of statistical inference, we conclude with an alternative recommendation.

Continuing with the example already described, suppose the analyst computes estimates of the path coefficients, p_{41} and p_{43}, by some method (not necessarily the OLS estimates recommended above). If these estimates—call them $\tilde{p}_{41}$ and $\tilde{p}_{43}$—are combined with sample correlations according to the normal equations, we have

$$r_{14}^* = \tilde{p}_{41} + \tilde{p}_{43}r_{13}$$
$$r_{24}^* = \tilde{p}_{41}r_{12} + \tilde{p}_{43}r_{23}$$
$$r_{34}^* = \tilde{p}_{41}r_{13} + \tilde{p}_{43}$$

where r_{hj} is an observed sample correlation and r_{hj}^* is the "implied" ("predicted" or "reproduced") correlation that would be observed if the overidentifying restriction(s) held exactly in the sample. Because of sampling error, implied and observed correlations will ordinarily not all be equal. Thus, we have a set of discrepancies,

$$d_{14} = r_{14} - r_{14}^*$$
$$d_{24} = r_{24} - r_{24}^*$$
$$d_{34} = r_{34} - r_{34}^*$$

If the OLS method of estimation were used, we would have

$d_{14} = d_{34} = 0$ but $d_{24} \neq 0$. If some other method were used we would still find one or more d's differing from zero. However, if the model holds true in the population, any such difference(s) should be "small," that is, no larger than one might reasonably expect as a consequence of sampling error alone.

It would seem plausible to use the set of d's in a formal statistical test against the null hypothesis which asserts the truth of the overidentifying restriction(s) of the model. However, this procedure is less convenient than the standard test described later, in the event that this test is available. Under some circumstances—though not in the case of the model used as an example here—the method of implied correlations may be recommended cautiously as a heuristic expedient, if an appropriate standard statistical test is not available. It may also be of use in the initial stage of specifying a model, where the investigator wishes to make an informal test of his ideas.

A similar approach to testing of overidentifying restrictions is the Simon–Blalock procedure. Consider any fully recursive model in which one or more paths are taken to be missing, that is, to have the value zero. It is then possible to deduce that certain simple and/or partial correlations will be zero. Blalock (1962–1963) has actually provided an exhaustive enumeration of the "predictions" for all possible four-variable models. The example given here appears in Blalock's enumeration as "Model E," and we find that in this model $\rho_{24.13} = 0$. The corresponding sample partial correlation ($r_{24.13}$) should, therefore, be close to zero. If it is not—if the difference from zero is too great to attribute to sampling error—we should be obliged to call into question the overidentifying restriction of this model. Blalock does not develop formal procedures of statistical inference for this kind of test. Again, we conclude that the proposed test is conceptually valid and is useful to the investigator who wants to be sure he understands the properties of his model. The research worker should not, however, rely on mere inspection of sample partial correlations. We recommend instead the standard statistical test described hereafter.

In our example, the issue as to the specification of the model is whether $p_{42} = 0$ or $p_{42} \neq 0$. In other words, we must decide as between the competing specifications of the x_4-equation:

$$x_4 = p_{43}x_3 + p_{41}x_1 + p_{4w}x_w$$

and

$$x_4 = p_{43}x_3 + p_{42}x_2 + p_{41}x_1 + p_{4w}x_w$$

We proceed on the latter specification and estimate by OLS the equation which includes p_{42}. In the usual routine for multiple regression we obtain as a by-product of our calculations the quantities necessary to compute the standard errors of our estimated coefficients (see, for example, the chapter on multiple regression in Walker and Lev, 1953, where the procedures are described for standardized variables). We may then form the ratio,

$$t = \hat{p}_{42}/\text{S.E.}(\hat{p}_{42})$$

and refer it to the t-distribution with the appropriate degrees of freedom. Roughly speaking, if one is working with a reasonably "large" sample, when $|t| \geq 2.0$, we may conclude with no more than 5% risk of error that the null hypothesis is false. In this event, we would reject the overidentifying restriction of the model and, presumably, respecify it to include a nonzero value of p_{42}.

In the case of failure to reject the null hypothesis—that is, if the t-ratio is not statistically significant—the situation is intrinsically ambiguous. Clearly, one is not obliged to *accept* the null hypothesis unless there is sufficient a priori reason to do so. It could happen, for example, that the true value of p_{42} is positive but small, so that our sample is just not large enough to detect the effect reliably. If our theory strongly suggests this is the case, we would do well to keep p_{42} in the equation despite the outcome of the test. In any event, it is good practice to publish standard errors of all coefficients, so that the reader of the research report may draw his own conclusion as well as have some idea of the precision of the estimates of coefficients. A good discussion of the issues raised by tests of this kind is given by Rao and Miller (1971); their discussion is presented in the context of a single-equation model, but it carries over to the problem of testing an overidentifying restriction on any equation of a recursive model.

This is not the place to develop the theory and techniques of statistical inference. What has been said can be reduced to a simple rule: If OLS regression is the appropriate method of estimation, the theory and techniques of statistical inference, as presented in the literature on the multiple regression model, should be drawn upon when making tests of overidentifying restrictions. (Perhaps we should note that such a test will not always involve the t-statistic for a single coefficient but may lead to the F-test for a whole set of coefficients.) This rule does not cover all cases, as will become apparent when we consider models for which OLS is not the appropriate method of estimation.

It is vital to keep the matter of tests of overidentifying restrictions in perspective. Valuable as such tests may be, they do not really bear upon what may be the most problematical issue in the specification of a recursive model, that is, the causal ordering of the variables. It is the gravest kind of fallacy to suppose that, from a number of competing models involving different causal orderings, one can select the true model by finding the one that comes closest to satisfying a particular test of overidentifying restrictions. (Examples of such a gross misunderstanding of the Simon–Blalock technique can be found, among other places, in the political science literature of the mid-1960s.) In fact, a test of the causal ordering of variables is beyond the capacity of any statistical method; or, in the words of Sir Ronald Fisher (1946), "if . . . we choose a group of social phenomena with no antecedent knowledge of the causation or absence of causation among them, then the calculation of correlation coefficients, total or partial, will not advance us a step toward evaluating the importance of the causes at work [p. 191]."

Exercise. *Show how to express all the correlations in terms of path coefficients in the model on page 44, using Sewall Wright's multiplication rule for reading a path diagram.*

Exercise. *Change the model on page 28 so that* p_{43} *is specified to be zero. Draw the revised path diagram. Obtain an expression for the overidentifying restriction. Indicate how one would estimate the coefficients in the revised model, supposing the overidentifying restriction was not called seriously into question.*

FURTHER READING

See Walker and Lev (1953, Chapter 13) for a treatment of multiple regression with standardized variables. Sociological studies using recursive models are numerous; for examples, see Blalock (1971, Chapter 7) and Duncan, Featherman, and Duncan (1972).

4

Structural Coefficients in Recursive Models

In Chapter 3 a four-variable recursive model was formulated in terms of standardized variables. That procedure has some advantages.

(*1*) Certain algebraic steps are simplified.
(*2*) Sewall Wright's rule for expressing correlations in terms of path coefficients can be applied without modification.
(*3*) Continuity is maintained with the earlier literature on path analysis and causal models in sociology.
(*4*) It shows how an investigator whose data are available only in the form of a correlation matrix can, nevertheless, make use of a clearly specified model in interpreting those correlations.

Despite these advantages (see also Wright, 1960), it would probably be salutary if research workers relinquished the habit of expressing variables in standard form. The main reason for this recommendation is that standardization tends to obscure the distinction between the structural coefficients of the model and the several variances and covariances that describe the joint distribution of the variables in a certain population.