

from Cook, T. D., & Campbell, D. T. (1979).  
*Quasi-experimentation: Design and analysis  
issues for field settings* (Ch. 7, pp. 295-309 only).  
Chicago: Rand-McNally.

---

---

## Inferring Cause from Passive Observation

---

---

*The section of this chapter entitled "The Causal Analysis of Concomitancies in Time Series" was written for this book by Melvin M. Mark.*

The previous chapters have dealt for the most part with inferring cause from manipulated changes which were arbitrarily and abruptly introduced at a specified time into a measurement and comparison setting. This enables researchers to make inferences regarding the effect of the intervention. In contrast, the methods of the present chapter try to infer causal processes based on observations of concomitancies and sequences as they occur in natural settings, without the advantage of deliberate manipulation and controls to rule out extraneous causal influences.

The methods discussed in this chapter are usually categorized under headings such as "correlational methods" or "nonexperimental methods." We have no quarrel with the latter title, but the former no longer seems appropriate. This is because correlations in the ordinary sense are what one looks for in experiments as well as in descriptive studies of the nonexperimental sort. In both cases, one is interested in discovering whether certain variables covary with others, irrespective of whether any of the variables is manipulated. Correlations in the technical statistical sense could be used to analyze data from experiments as well as from noninvasive observational studies. The *t* test of the difference between means, the *F* test from analysis of variance or covariance, and the common multiple regression approaches are all part of the same general linear statistical model. The crucial difference between the methods discussed in this chapter and those used in experimentation is not in the choice of statistics to analyze the data. Indeed, even if only graphic or tabular means of analysis and presentation are used, these are just as applicable to the experiment as to descriptive or observational studies.

Traditionally, the term "correlational study" refers to a purely descriptive or observational study where the data are statistically analyzed. Since the terms "description" and "observation" could also apply to the data from experiments, we propose for the sake of clarity to add to them the adjective "passive," as in the title of this chapter. While we are not likely to change well-established usage,

we propose for our own use the contrasting phrases "experimental studies" and "passive-observational studies." To shorten the latter to "observational studies" alone would invite confusion with the similar distinction between "experimental" and "observational" studies in the statistics literature, under which all nonrandomized experiments—including quasi-experiments—may be classified as observational.

Passive-observational approaches may be employed for the purpose of causal inference. We have briefly alluded to this possibility in chapter 1, even while holding up deliberate, intrusive experimentation as the epitome of the meaning of causation. In chapters 5 and 6 on time series we applied experimental modes of analysis even when the "treatment" was a natural disaster rather than a planned intervention. But in this and other similar cases, we also specified that the event being evaluated had to be abrupt and precisely dated, and not a reaction to prior change in the level of the indicator. Given these restrictions, such a passive observational setting might justifiably be called a "natural experiment" since it approaches a deliberate experiment in form—a form which facilitates causal inference. In regard to the nonequivalent control group design, we could tolerate the use of the methods of chapters 3 and 4 for "treatments," such as attending a particular training program, even where the program was a permanent institution and where the researcher did not manipulate anything. However, in those chapters we added the caveat that it was crucial to render rival hypotheses implausible, such as selection-maturation interactions, and that there should be structural aspects of the setting which reduce the equivocality of causal inference normally accompanying correlational data. The settings in this sense should be quasi-experimental, substituting natural controls for those that would be introduced in deliberate experimentation. In the present chapter the goal of achieving causal inference from passive observational data is extended to include observations from settings which lack even such quasi-experimental structural controls.

In aspiring to causal inference from passive-observational data, it is essential to distinguish sharply between this goal and other uses of the same type of correlational evidence, such as forecasting. For purely forecasting purposes, it does not matter whether a predictor works because it is a symptom or a cause. For example, your goal may be simply to predict who will finish high school. In that case, entering the Head Start experience into a predictive equation as a negative predictor which reduces the likelihood of graduation may be efficient even if the Head Start experience improved the chances of high school graduation. This is because receiving Head Start training is also evidence of massive environmental disadvantages which work against completing high school and which may be only slightly offset by the training received in Head Start. In the same vein, while psychotherapy probably reduces a depressed person's likelihood of suicide, for forecasting purposes it is probably the case that the more psychotherapy one has received, the greater is the likelihood of suicide.

This chapter is *not* concerned with presenting forecasting techniques. They enter into discussion here only insofar as well-established modes of analysis developed for forecasting are still mistakenly used for causal inference. This confusion is all the more saddening because, in the first book on the subject, Blalock (1961) clearly distinguished between "prediction" and "causal inference." Moreover,

within the literature on regression techniques, one can find an explicit distinction between "predictive regression" (i.e., forecasting) and "structural regression" (i.e., causal inference or something close to it). Rogosa (1978) is able to trace such a usage back to Wald (1940) and Tukey (1954). Wold (1956) has long made a similar distinction. But in most applied statistical analyses in the social sciences this distinction is ignored. Multiple regression beta weights—for a program treatment (coded as a dummy variable, 1 for treated, 0 for untreated) or for some other antecedent variable—that are quite appropriate for forecasting purposes are mistakenly reported in causal language. For forecasting purposes, it makes no difference what the true causal path is from one variable to another, and there is no need to distinguish between symptoms and causes. Nor does it create a problem if the presumed causal variable is a complex composite of which only a part produces the correlation with the outcome variable. However, for structural regression, both error and independent factors in the covariates (the "independent variables") can produce profoundly misleading results, neglect of which can lead to *reversing* the sign of the beta weight, for example making the Head Start experience seem harmful. For predictive regression, leaving out a relevant covariate may produce a forecast less precise than it might otherwise have been but does not lead to any misstatement of the forecasting relevance of the covariates employed. For structural regression, such an omission or specification error may again lead to erroneous estimates of both the magnitudes and signs of the causal paths among the variables at hand. Thus the distinction between *causal inference* from observational data and *forecasting* from the same data is a very significant one with many concrete methodological implications. It is on the causal inference agenda that we focus in this chapter. Many readers are already fluent in the algebra and computer programs for forecasting. We ask you in particular to attend to this difference.

Behind the prevalent confusion of the two purposes may be the well-established terminological usage of calling the covariates in multiple regression "independent variables" even though many of them cannot be "independently varied" even with an all-out massing of the powers of government. How John Stuart Mill's precious and appropriate usage for deliberate manipulation (see chapter 1) could come to be so misapplied is a question for some historian of science to answer. Doubtless it occurred in that period from Mill to Russell when Hume's denial of any distinction between causation and correlation dominated the theory of science. Perhaps Karl Pearson himself was to blame, being at the same time a positivist philosopher of science, an evolutionary biologist who fought the introduction of Mendel's theory with its unobserved causal genes, and a major contributor to the development of multiple regression statistics.

In this chapter we present three or four common modes of quasi-experimental analysis that will be encountered in the research literature. For the statistically sophisticated reader, these approaches are all redundant alternative presentations of a common mathematical structure presented more elegantly in chapter 4. Yet we believe it is important to discuss the crucial issues in each of these forms of presentation, since it is in these forms that causal claims from field research will most commonly be encountered. The four sections that follow can be read in any order. Their relative difficulty will vary depending upon the statistical background of the reader. The section on "treatment as a dummy variable in a regression

analysis" is particularly condensed and requires prior familiarity with regression and adjustments to third variables. The path analysis section attempts to provide some introductory understanding to those not previously familiar with the approach, but omits any introduction on how to do such analyses. The section on the cross-lagged panel correlation gives considerable detail on the past history of the method, since we have also taken the occasion to update our view of its historical development. The final section, on drawing causal inferences from the relationship of two time series, is only a brief sketch of what is a voluminous and growing literature.

### TREATMENT AS A "DUMMY VARIABLE" IN A REGRESSION ANALYSIS

In the actual practice of applied social science, including program evaluation, the most common mode of causal inference, the most common quasi-experimental design, is to represent the treatment as a "dummy variable" (treated = 1, untreated = 0) in a regression analysis. Since the treated group differs in composition from the untreated, other variables are entered into the analysis in the expectation that they will model the selection process whereby different kinds of persons came to be in different groups, or will provide an acceptable causal model of the dependent variable. A treatment effect is inferred if there is a statistically significant regression coefficient (beta weight) relating the dummy variable to the dependent variable, after adjusting for the effect of the covariates introduced to try and correct for selection. Unfortunately, those who use this widespread practice make the error of employing forecasting techniques for causal inference.

Our main analysis of the mistaken assumptions underlying this approach is presented in Reichardt's chapter 4. Here we discuss a commonsense illustration stated in the terms of regression analysis. For the older generation of readers, we will use the language of partial correlation, asking the (misleading) question, e.g., What is the correlation between the treatment (T) and achievement (A) when we have "partialed out" the effects of social class (S), family income (F), education (E), etc. (i.e., what is  $r_{AT \cdot SFE}$ )? For those with more modern training in multivariate statistics, we will speak in terms of adjusted regression coefficients, or betas,  $\beta_{AT \cdot SFE}$ . We are addressing this illustration, remember, not to the best methodology articles and texts in applied social science, but to a common practice which is all the more entrenched because of handy computer programs that generate "impressive" output.

Table 7.1 presents a hypothetical set of correlations based upon a model situation in which most users of multivariate regression techniques would expect these techniques to work. In this hypothetical simulation, let us imagine that the population is from middle management. The pretest (Pr), which we will not use for most of this simulation, and posttest (Po) are similar measures of executive aptitude. These scores have 80% of their variances determined by the same underlying "social advantage" factor. The experimental treatment (T) could be a year in a graduate school of management, intentionally awarded to the most promising, and on the whole, to those highest on a latent social advantage factor. In our hypothet-

ical example, this factor accounts for half of the variance in assignment to treatment and is the sole source of the .505 correlation between treatment and pretest. Since T is a dichotomous variable, this correlation is lower than it would otherwise be. In terms of shared components of variance, it corresponds to the .632's found between the pretest and the covariate in Table 7.1. Similarly, the .399's in that table are reduced by T's dichotomous character from the .500's that would be obtained if both variables were continuous. The correlation between the treatment and pretest shows that those who received the treatment were already ahead of the others in executive aptitude. For our hypothetical case, the treatment has had no real effect on executive ability. Thus the correlation of the treatment with the pretest is identical to its correlation with the posttest and is again due to the shared social advantage factor as reflected in selection of candidates for the program, rather than the beneficial effects of the program.

	(Pr)	Po	T	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>
Po	(.800)					
T	(.505)	.505				
C <sub>1</sub>	(.632)	.632	.399			
C <sub>2</sub>	(.632)	.632	.399	.500		
C <sub>3</sub>	(.632)	.632	.399	.500	.500	

Table 7.1. Hypothetical intercorrelations among variables for regression adjustment demonstration.

Shown also in Table 7.1 are three excellent covariates tapping the same latent social advantage factor which determines 50% of the variance of each. To make these illustrations more concrete, imagine that C<sub>1</sub> is father's income, C<sub>2</sub> mother's educational level, and C<sub>3</sub> the socioeconomic level of the candidate's home neighborhood. Let us now use these covariates to adjust the relationship between the treatment and the posttest measure of executive ability in the hope of removing the effects of the social advantage factor that determined who received the treatment—a treatment that we simulated to be ineffective. Table 7.2 shows such adjustments for a version of our hypothetical illustration in which no pretest is available. The first row gives the starting values, with no covariates employed. The adjusted mean differences are stated in terms of standard score units for the pooled population of those who received the treatment and those who did not. These differences directly correspond to the partial correlations and beta weights also presented in the table.

Using one covariate substantially reduces the apparent effect of the program. Each covariate added helps reduce the apparent effect still more. With three of these unusually powerful covariates, the adjusted mean difference is down to .33, the partial correlation  $r_{POT \cdot C_1 C_2 C_3} = .230$ , the regression coefficient  $\beta_{POT \cdot C_1 C_2 C_3} = .167$ . But these values are still of a magnitude that would be mistakenly reported as impressive results for the training program. If the study were based on 500 cases, half receiving the training, these differences would be significant beyond the  $p < .0001$  level. Even if we had ten such marvelous covariates measuring social advantage, with 500 cases the adjusted mean difference would still be significant at the posttest ( $p < .02$ ) level. Thus even in this apparently optimal situation for the use of regression adjustments, they fail. The underadjusted residuals left could be mistakenly reported as treatment effects. The only way such covariates can provide an adequate adjustment is for one or more of them to be a perfectly reliable and valid measure of the latent common factor. Such covariates simply do not exist, and most covariates available have much weaker relevance than do those of this hypothetical illustration.

Number of covariates	Adjusted regression coefficient $\beta_{POT \cdot C_1 \dots C_n}$	Partial correlation $r_{POT \cdot C_1 \dots C_n}$	Adjusted mean difference on posttest
0	(.505)	(.505)	(1.01)
1	.301	.356	.60
2	.214	.278	.42
3	.167	.230	.33
10	.065	.105	.13

**Table 7.2.** Regression adjustments on the matrix of Table 7.1 (see text).

What if we do have the pretest and adjust on it? The adjusted mean difference would be .271,  $\beta_{POT \cdot Pr} = .136$ ,  $r_{POT \cdot Pr} = .195$ , all of which spuriously suggest a treatment effect. What if we use both the pretest and the three other covariates? The adjusted mean difference becomes .175,  $\beta_{POT \cdot Pr C_1 C_2 C_3} = .111$ ,  $r_{POT \cdot Pr C_1 C_2 C_3} = .136$ . All of these, though small, still have a magnitude frequently reported as treatment effects. It is especially worth noting that our illustration has assumed but one underlying latent factor. If the factorial pattern is more complex, both multivariate regression adjustments and partial correlation are equally inappropriate. Only by a chance combination of circumstances can they give the correct adjustment.

Two striking reanalyses of previous multivariate regression estimates of program effects have recently been published that questioned the initial analyses which found compensatory educational efforts *harmful*. Magidson (1977) reanalyzed data designed to evaluate Head Start. While his alternative analysis has come under attack (Bentler and Woodward, 1978; Magidson, 1978), it should be noted that the quarrel is about whether a plausible alternative analysis shows statistically significant beneficial effects. In both Magidson (1977) and Bentler and

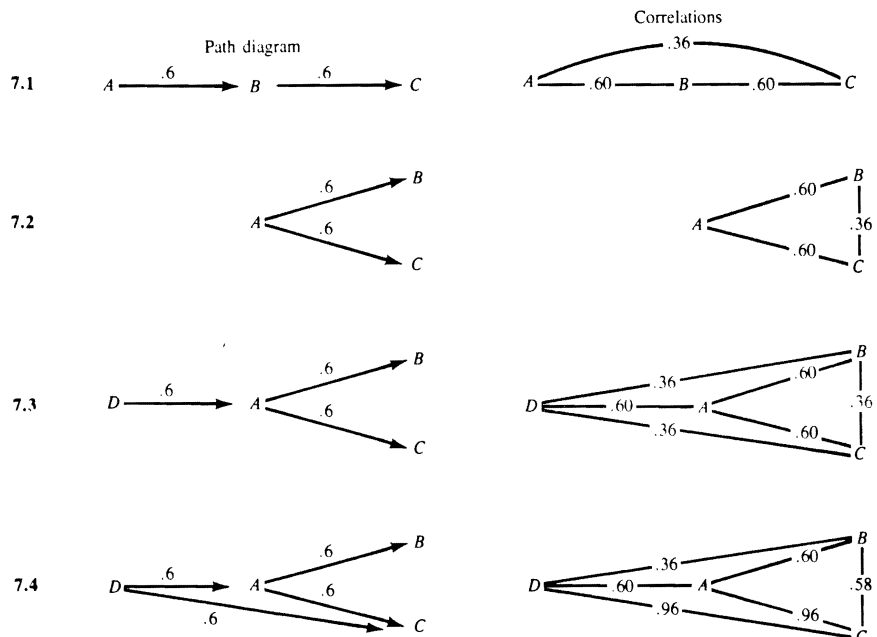
Woodward (1978) the sign for the treatment effect is positive and is opposite to that obtained from the earlier multivariate regression analyses which led to conclusions about harmful effects. Director (1979) called attention to the misleading conclusions of regression analyses of Job Corps training, which again have found such training a liability. Having in one study three waves of measurement with no training between the first two, Director was able to apply the standard regression approach to a "dry run" estimate of the effect of the training program before it was ever implemented. The approach resulted in significantly harmful effects from the nonexistent training. Director's alternative analyses removed the appearance of harmful effects in all cases, but failed to find significantly positive effects from such programs. The point to be emphasized here is that the use of standard regression approaches often results in unwarranted conclusions of harmful treatment effects when "compensatory" treatments are evaluated where the persons receiving the treatment do less well than controls at the pretest. Similarly, they often lead to unwarranted conclusions of beneficial impact when programs are evaluated where the initially more fortunate group has greater exposure to the treatment, as was the case in the hypothetical example above.

## CAUSAL MODELING BY PATH ANALYSIS

In the last decade techniques for causal inference from nonexperimental data have emerged as dominant within sociology, economics, and political science under such names as causal models, path analysis, and structural equation models. The dominance of these methods can be easily verified by consulting contemporary theory-related empirical studies in sociology as well as methodological journals such as *Sociological Methods and Research* and *Social Science Research*. Indications are that this dominance will increase rather than decrease during the next few years. This is not the place to begin an education in path analysis, nor are we appropriate teachers. Blalock (1961, 1971), Duncan (1975) and Heise (1975) provide good starting points, while the Goldberger and Duncan (1973) text is more advanced. Nonetheless, it seems useful to present some of the basic ideas at their simplest level and to use this important language of explanation and analysis to restate some of the concerns that have been raised throughout this book.

In practice, most of the detailed methodology of path analysis deals with estimating path coefficients and paths from correlation coefficients computed on observational data. Didactically, however, it is simpler to start as though the path coefficients were known and generate the correlation coefficients that would result from them. This is particularly true when illustrating methodological problems. No matter what the complexity of an assumed path model, one can always generate the relevant correlation coefficients. This is true even in cases in which it would be impossible to do the reverse and deduce path coefficients from the manifest correlational data.

Let us start with a very simple three-variable model, Figure 7.1, where *A* causes *B* and *B* in turn causes *C*, there being no direct causal connection from *A* to *C*: i.e.,  $A \rightarrow B \rightarrow C$ . The two path coefficients generate the three correlations as follows: First, the path coefficient itself gives the correlation between two directly connected variables (except as this value is modified by a correlation mediated



Figures 7.1, 7.2, 7.3, and 7.4. Illustrative causal path diagrams and the corresponding correlation coefficients.

indirectly by common causes which the two variables share, as in subsequent examples). Thus  $r_{AB} = .60$  and  $r_{BC} = .60$ . Second, mediated causal relations generate correlation components equal to the product of the path coefficients of the links that connect them. Thus  $r_{AC} = .6 \times .6 = .36$ . Note that this same pattern of correlations would be generated by these alternative paths:  $A \leftarrow B \leftarrow C$  and  $A \leftarrow B \rightarrow C$ . Like the  $A \rightarrow B \rightarrow C$  model under test, these other models imply no path from  $A$  to  $C$ . Thus if one found that  $r_{AC} = .36$ , this would imply no causal path from  $A$  to  $C$ , but it would not allow one to differentiate among the three causal models unless there was background information or trusted theory which allowed one to rule out some of the models on a priori grounds. (Regression coefficients can be used to test simple models like this. Assuming the model is correct,  $\beta_{AC \cdot B} = 0$ .)

Figure 7.2 shows another three-variable case, to which we will add more variables to illustrate further computational rules. In Figure 7.3  $B$  and  $C$  now share a remote common cause  $D$  as well as the direct common cause  $A$ . But this does not add to the correlation between  $B$  and  $C$  since the routes from  $D$  to  $B$  and from  $D$  to  $C$  both have to go through  $A$ . Therefore  $A$  already transmits all of the correlation-

generating causal efficacy of  $D$ , and  $D$ 's addition does not change  $r_{BC}$ . However, if we add a direct causal path from  $D$  to  $C$  as in Figure 7.4,  $r_{BC}$  is changed. The new path creates a supplementary channel for  $D$ 's influence on  $r_{BC}$ . To the .36 correlation points contributed by the fact that both  $B$  and  $C$  share  $A$  as a cause, mediated through the loop  $BAC$ , are added the .22 correlation points generated by  $D$  as a common cause, mediated through the new loop  $BADC$  ( $.6 \times .6 \times .6 = .216$ ) summing to the new value,  $r_{BC} = .58$ . This new path also changes two of the correlations corresponding to direct paths. The correlation  $r_{AC}$  becomes .96, composed of the direct path .60 plus the influence of their shared cause  $D$ , via loop  $ADC$ ,  $.6 \times .6 = .36$ . The value for  $r_{DC}$  also becomes .96 via  $DAC$  as well as  $DC$ .

We have moved in the easy deductive way from known path coefficients to resultant correlation coefficients. We will not attempt here to teach how to move in the inductive direction from the correlation coefficients (plus the assumed paths) to estimates of the path coefficients. For all of the cases of Figures 7.1 through 7.4 it can readily be done. Since in each case there are more correlation coefficients than hypothesized paths, the models are "over-identified." This means that there are enough degrees of freedom not only to estimate the paths but also to test the goodness of fit of the model. Where the number of coefficients and the number of paths are equal, the model is just-identified, and one can estimate coefficients but not test the fit. Where, as in later examples, there are more paths than correlation coefficients, the path coefficient cannot be estimated. Even in such cases, however, drawing up path diagrams using the best available background information, the most trusted theories, and making plausible guesses as to path coefficients can be of fundamental help in the theory-relevant interpretation of correlational data.

In discussing the examples of Figures 7.1 through 7.4, we have not specified how measurement errors (unreliability) and reliable irrelevant components of variables (invalidity) enter into consideration. However, one way of conceptualizing the totality of such unique or unshared components is already implicit in these path models. Figure 7.5 provides a more explicit version of Figure 7.1 to clarify this conceptualization. In path analysis, the variables have all been "standardized," with the variance of each variable being transformed to 1.0. The causal paths impinging upon a variable usually leave much of that variance unexplained. The unexplained residual can be attributed to unique components, spelled out in

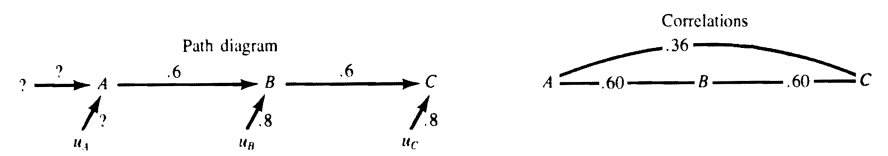


Figure 7.5. A version of Figure 7.1 with the unique components ( $u$ ) of measured variables made explicit. (Lower case letters represent unobserved variables.)

Figure 7.5 using  $u$  for unique components, with the letter in lower case to indicate that it is an unobserved variable. (For  $B$  and  $C$ , values for the path from  $u$  have been set at .8 so that the sum of the squares of impinging paths will equal 1.00 ( $.8^2 + .6^2$ ) and no remote sources of correlation complicate the picture. For  $A$ , which is unexplained in this model, the path  $u_A \rightarrow A$  is indeterminate, and could be set from .00 to 1.00 without affecting the rest of the model.) As noted for variable  $D$  in Figure 7.3, a cause which enters a network through only one variable can be omitted without affecting the paths or correlations among the other variables. Especially where such causes are unobserved variables, it simplifies the presentation to leave them out (as in Figure 7.1). They do not affect the computational aspects of the method either for deducing correlations from path coefficients or for inferring path coefficients from observed correlation coefficients. However, putting them in (as we have in Figure 7.5) does make more explicit the nature of the conceptual model. It also helps explain why in the most advanced presentations of causal modeling this is *not* the recommended approach for representing error and unique variance in variables, even though it is the one approach implicit in most of the older and many of the current research applications of path analysis.

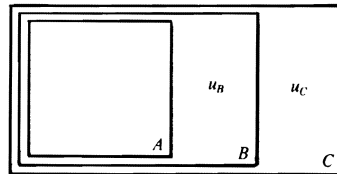


Figure 7.6. Shared variance overlap diagram corresponding to Figures 7.1 and 7.5.

In the model of Figure 7.5 (and Figure 7.1), the unique components of  $B$ ,  $u_B$ , are not irrelevant to  $C$ , but become a causally efficacious part of  $B$  impinging on  $C$  fully as much as those components deriving from  $A$ . The overlapping-area diagram, Figure 7.6, using an older approach to understanding correlation, may help make this clearer for some readers. All of the variance of  $A$  is contained in (i.e., causally effective in)  $B$ , and all of the variance of  $B$  is contained in  $C$ . The correlation between  $A$  and  $B$  is less than perfect only because of  $u_B$ , the unique components of  $B$ . Similarly, it is  $u_C$ , not  $u_B$ , which keeps  $r_{BC}$  down to .60 instead of being 1.00. But this is a very implausible model. In most situations, we know that our variables are measured imperfectly, with both irrelevant systematic sources of variance as well as haphazard error of which "methods factors" (Campbell and Fiske, 1959) are only one type. Except in those rare instances in which a measurement is itself directly used in a social decision process (which we discuss in regard to the regression-discontinuity design in chapter 3), it is not reasonable to regard these incidental components of the measure's variance as being causally efficacious. Furthermore, we know that in practice correlations are apt to be atten-

uated by the unique components of the "cause" fully as much as by the unique components of the "effect." The model of Figures 7.1, 7.5, and 7.6 make the mistake, discussed in previous chapters, of assigning all the error to the "dependent" variable, treating an "independent" variable as though it were a perfect measure.

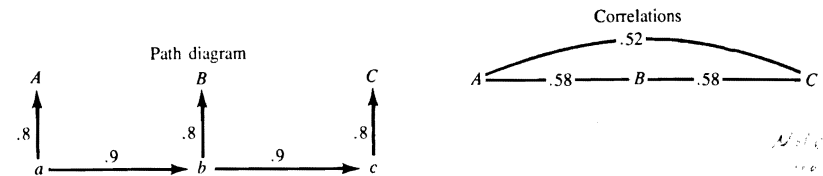


Figure 7.7. Path model with the preferred representation of unique variance in measured variables. (Capital letters represent measured variables, lower case are unobserved.)

Figure 7.7 shows a much more appropriate way of modeling the unique variance in measured variables. Here the causally efficacious relations are among the unmeasured or latent variables  $a$ ,  $b$ , and  $c$ , rather than the observed  $A$ ,  $B$ , and  $C$ . (Since much of the unique variance that attenuated relations in Figures 7.1 and 7.5 is now moved up out of the causal path, the path coefficients relating  $a$ ,  $b$ , and  $c$  have been set at larger values. To avoid confusion, paths from unique components have been omitted.)

While this model is no doubt more realistic, it is much less "convenient." There are five paths to estimate from only three observed correlation coefficients; thus the model is seriously underidentified. Only by making further assumptions could one begin to estimate causal paths. For example, one might assume that the reliability-validity of each of the variables was the same, so that  $a \rightarrow A = b \rightarrow B = c \rightarrow C$ . This reduces to three the number of path coefficients to be estimated, and the model becomes just-identified. While this is an unreasonable assumption, almost certain to be wrong unless  $A$ ,  $B$ , and  $C$  are three forms of the same intelligence test, it is clearly a more reasonable approximation than to assume the model of Figures 7.1, 7.5, and 7.6 where  $A$  is measured without uniqueness, and  $B$ 's uniqueness from  $A$  is causally efficacious in  $C$ .

If the true but unknown model of causal relationships takes the form of Figure 7.7, but we misinterpret it and test the model in Figure 7.1, we are led to the seriously mistaken conclusion that there must be a direct causal path between  $A$  and  $C$ , inasmuch as  $\beta_{AC \cdot B} = .27$  rather than zero.

Practical approaches are being developed regarding the problem of unique variance in measured variables. These involve causal models for unobserved variables, and in our judgment this development is essential if path analysis and structural equation modeling is to minimize misleading conclusions. The work of Jöreskog, Sörbom, and Goldberger (Jöreskog, 1970, 1973, 1974, 1977; Jöreskog and Sörbom, 1976, 1978; Jöreskog and Goldberger, 1975; Goldberger, 1972, 1973; Long, 1976) seems particularly important. Such models in practice will

often involve assumptions that one knows are oversimplifications. Nonetheless, such assumptions will almost always be more appropriate than those made in the simpler form of path analysis and in ordinary regression analysis, namely, that the causal variables and other covariates have been measured without unique irrelevant variance. Note that in the vigorous exchange already cited between Magidson (1977, 1978) and Bentler and Woodward (1978) on Head Start reanalyses, both are employing the Jöreskog-Sörbom techniques.

We can continue this introductory presentation of path analysis by using it to model one of the imperfect compensatory examples referred to in the first section of this chapter. In school systems covering a range of educational advantages in the homes of its pupils, the observed correlation between Head Start attendance (*HS*) and first-grade achievement (*Ach*) will be negative. If we use  $HS \rightarrow Ach$  as a complete causal model, we would conclude that the path coefficient is negative and that Head Start has a harmful effect on school achievement. But such a simple model is guilty of a specification error, since it has omitted at least one important causal variable that influences achievement. We can call this the educational advantage (*EA*) associated with home and neighborhood environments. Specification error is an important problem for causal modeling but we will not elaborate on it here. It is enough to say that valid causal modeling is dependent on functionally complete specification and that whenever path coefficients are taken seriously as a research product complete specification has to be assumed. Omitting a relevant cause not only can produce an erroneous estimate of the size of the path coefficients among the variables attended to, it can also lead to the sign of the path coefficient being wrong, mistaking negative causation for positive and harmful effects for beneficial ones. In the very simple Head Start case above, this type of omission occurred.

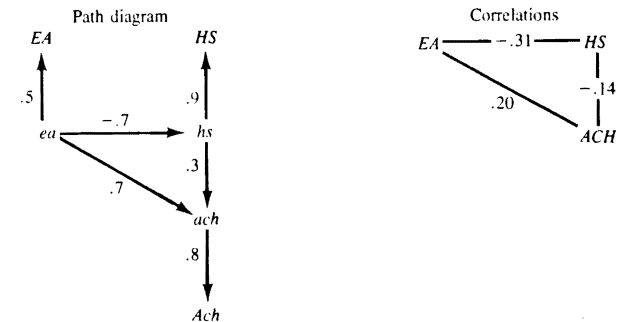


**Figure 7.8.** A model for educational advantage (*EA*), Head Start (*HS*), and school achievement (*Ach*), lacking proper representation of the unique components of the measured variables.

Figure 7.8 illustrates a somewhat more complete model. Head Start has been offered to those from the least advantaged homes, hence the negative path coefficient between *EA* and *HS*. Moreover, we know from past research that children from the most advantaged homes do better in first-grade achievement, hence the positive path coefficient between *EA* and *Ach*. Assume now that the Head Start experience improves first-grade achievement, resulting in the positive coefficient from *HS* to *Ach*. The resultant correlation between Head Start and achievement is

made up of this direct causal component .30, and the indirect component due to the common cause, educational advantage, shared by *HS* and *Ach*. This indirect component ( $.7 \times -.7$ ) is  $-.49$ , the net correlation between *HS* and *Ach* is therefore  $-.19$ . However, if such a model were correct, one could successfully control for the effect of educational advantage by a regression adjustment,  $\beta_{Ach \cdot HS \cdot EA} = .30$ , showing positive benefits from Head Start in accordance with the model.

But the model of Figure 7.8 has not represented the unique components in the measured variables adequately. If we try to do this, an example such as Figure 7.9 results. (In assigning the path coefficient between *ea* and *EA*, we have been guided by our impression of the validity of such measures. For  $hs \rightarrow HS$  we have used .9 rather than 1.0, in awareness of errors in school records and cases of near zero attendance.)



**Figure 7.9** A preferred model for representing unique components in measured educational advantage, Head Start, and school achievement.

In this case, where unique variance in measured variables is more appropriately modeled, the regression adjustment does *not* correct the direction of effect. Indeed,  $\beta_{Ach \cdot EA} = -.08$ , a value that still suggests small but harmful effects in the same direction as the simple correlation between *HS* and *Ach*. Similar models could of course be supplied for other illustrations, such as the probably higher suicide rate for those who have received more psychotherapy, or the lower calculus grades for those with more tutoring, and other instances of valid compensatory efforts that do not offset initial disadvantages.

As a contribution to competent causal theory building in complex real-world settings, path analysis can perform two quite different functions: theoretical clarification and estimation of specific causal impacts. As a tool for theoretical clarification, it has our strong approval. Starting from one's theoretical, empirical, and commonsense knowledge of a problem, one maps all of the latent variables one believes to be present and the probable causal links among them, indicating which paths are positive and which negative. One adds in the measured variables already available and those that might be obtained. While the path-analytic language uses

linear relationships, at this stage one can even note links that are believed to be nonmonotonically curvilinear (where, for example,  $A$ 's maximum effect in increasing  $B$  is at an intermediate value of  $A$ ). By adding in estimated path coefficients, one can derive predictions about the size and direction of the correlation coefficients that one would expect between available measured variables. Such derivations often produce surprises such as *negative* correlations connecting two variables linked by a *positive* causal path, thus showing how deceptive common habits of drawing causal inferences from correlational evidence can be. (In Figures 7.8 and 7.9, the relation between Head Start and achievement provides a simple example of such an outcome.) Used in this way, drawing up causal pathway maps as complete and complex as the real-world setting itself can be an enormously valuable preparation for field work, whether of a quantitative or qualitative nature. And with regard to the specific issues of this book, path analysis used this way increases awareness of the equivocality of causal inference from passive observational data and, by contrast, of the great advantages in reducing equivocality of experimental control, randomization, comparable comparison groups, and pretests similar to posttests. It is a flexible language in which the precautionary messages about third-variable causation and the other dangers of inferring causation from correlation can be clearly expressed.

This use of path modeling for making explicit what one knows about a real-world causal nexus is rare, but it is no doubt sometimes used in unpublished stages of research. However, it is antithetical in spirit to the commoner goal of producing estimates of specific causal path coefficients. Such estimates require greatly reduced complexity, wholesale elimination of plausible causal connections, and usually the elimination of the distinction between measured variables and the causally effective variance they contain. The estimates of specific causal paths may occasionally be plausible and valid, but the pressure to come up with a model permitting their estimation results in omissions which render most of these conclusions suspect. In terms of the concerns of this book, the omissions of error and unique variance in the causal variables are particularly serious.

These cautionary notes are of course not ours alone. They are to be found in Blalock's (1961) book and its successors. It is fitting to close this section with two quotations from Duncan's (1975) recent text:

We distinguish sharply between (1) statistical description, involving summary measures of the joint distributions of observed variables, which may serve the useful purpose of data reduction, and (2) statistical methods applied to the problem of estimating coefficients in a *structural model* (as distinct from a "statistical model") and testing hypotheses about that model. One can do a passably good job of the former without knowing much about the subject matter (witness the large number of specialists in "multivariate data analysis" who have no particular interest in any field). But one cannot even get started on the latter task without a firm grasp of the relevant scientific theory, because the starting point is, precisely, the model and not the statistical methods (pp. 5-6).

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. 50).

## CROSS-LAGGED PANEL CORRELATIONS

The cross-lagged panel correlation technique is much less widely used than path analysis or structural equation modeling. It is restricted to a much narrower set of passive observational settings, where two or more variables have been measured on a number of persons (or other units) on two or more occasions. The method was initially proposed independently of the path analysis tradition, and its preferred measure of cause has been the comparison of correlation coefficients between measured variables rather than latent path coefficients. However, when its assumptions are made explicit, they can be presented in path analysis form. We give this technique disproportionate attention here, including commentary on its historical development, because the predecessors of this book (Campbell and Stanley, 1963, 1966; Cook and Campbell, 1976) have been a major channel for disseminating information about cross-lagged panel correlations. While these earlier presentations have treated the technique with skeptical advocacy, this presentation carries still more skepticism and even less advocacy.

Beginning some 30 years ago, Lazarsfeld (e.g., 1947, 1948, 1972; Lipset, Lazarsfeld, Barton and Linz, 1954. See Lazarsfeld, 1978, for a history) argued that repeated measurement of the same two variables potentially should provide information about the direction of any causal asymmetries between them. His "16-fold table" (two dichotomous variables at two successive times generating 16 cells) probably finds its best modern descendant in Goodman (1973, 1978). In an effort to convert the 16-fold table for use with continuous variables and to relate it to the use of lagged correlations in economics, Campbell (1963; Campbell and Stanley, 1963) devised the cross-lagged panel correlation method. Working independently, and with more impressive illustrations, Pelz and Andrews (1964) developed essentially the same model. Kenny has contributed many further developments (1973, 1975b, in press).

Figure 7.10 presents the set of terms and images we will use in this presentation. The letters  $A$  and  $B$  represent two variables, each measured at two times—1 and 2—on the same persons (i.e., a "longitudinal" study or, in social survey terms, a "panel study"). The initial question at issue was whether  $A$  is a stronger cause of  $B$  than  $B$  is of  $A$ ? A careful note should be made of the question, and it should be differentiated from the concern in previous chapters which can best be summarized as: Is  $A$  a cause of  $B$ ? The synchronous correlations,  $r_{A_1B_1}$  and  $r_{A_2B_2}$ , and the test-retest correlations,  $r_{A_1A_2}$  and  $r_{B_1B_2}$ , provide the interpretative framework for the two cross-lagged correlations,  $r_{A_1B_2}$  and  $r_{B_1A_2}$ . If  $A$  is a stronger cause of  $B$  than  $B$  is of  $A$ , and if there are real storage processes (such as bank accounts and memories) spreading out the causation in time, then it might be