*Journal of Educational Statistics*
*Summer 1987, Vol. 12, No. 2, pp. 101–128*

# As Others See Us: A Case Study in Path Analysis

**D. A. Freedman**
*University of California at Berkeley*

Key words: *path analysis, path models, structural equation models, regression, correlation, causation*

*In 1967, Blau and Duncan proposed a path model for education and stratification. This is one of the most influential applications of statistical modeling technique to social data. There is recent use of the same technique in Hope's (1984) comparative study of Scotland and the United States,* As Others See Us: Schooling and Social Mobility in Scotland and the United States. *A review of path analysis is offered here, with Hope's model used as an example, the object being to suggest the limits of the method in analyzing complex phenomena.*

Argumentation cannot suffice for discovery of any new work, since the subtlety of nature is greater many times than the subtlety of argument.

Francis Bacon

## Introduction

The path model for education and social stratification proposed by Blau and Duncan (1967) was one of the first applications of that method in the social sciences, and certainly the most influential. It is often cited as a great success story: see, for example, Adams, Smelser, and Treiman (1982, p. 46). And it set the pattern for much subsequent research.

Indeed, path models are now widely used in the social sciences, to disentangle complex cause-and-effect relationships. Despite their popularity, I do not believe they have in fact created much new understanding of the

I would like to thank Persi Diaconis, Otis Dudley Duncan, Arthur Goldberger, Erich Lehmann, Thomas Rothenberg, Juliet Shaffer, and Amos Tversky for many useful discussions—without implying their agreement (or disagreement) on the issues. Neil Henry was a wonderful referee.

The quotations from Keith Hope (1984), and Figure 6, are reprinted with permission from *As Others See Us: Schooling and Social Mobility in Scotland and the United States,* copyright © 1984, Cambridge University Press. The research in this paper was partially supported by NSF Grant DMS 86–01634.

phenomena they are intended to illuminate. On the whole, they may divert attention from the real issues, by purporting to do what cannot be done—given the limits on our knowledge of the underlying processes.

One problem noticeable to a statistician is that investigators do not pay attention to the stochastic assumptions behind the models. It does not seem possible to derive these assumptions from current theory, nor are they easily validated empirically on a case-by-case basis. Also, the sheer technical complexity of the method tends to overwhelm critical judgement.

I have no magical solution to offer as a replacement. On the other hand, repeating well-worn errors for lack of anything better to do can hardly be the right course of action. If I am right, it is better to abandon a faulty research paradigm and go back to the drawing boards.

Blau and Duncan (1967) may seem like ancient history, so I will illustrate the argument by reference to Hope (1984). This is a comparative study of education and stratification in Scotland and the United States. Different chapters of the book take up different substantive issues (which are all interesting); it seems fair to look only at the first chapter. I ask Hope's pardon for using his work this way, but it provides a convenient example of the path-analytic paradigm in action.

At bottom, my critique is pretty simple-minded: Nobody pays much attention to the assumptions, and the technology tends to overwhelm common sense. Since the point is such an elementary one, the argument should start close to the beginning. After saying in statistical language what path models assume and what they do, I will outline Hope's model, and point to the absence of anything connecting it to reality—other than the conventions of the paradigm and his desire. Finally, some of the metamodeling literature will be reviewed.

## A Research Fable

Later sections will review the foundations of regression models in formal detail. This section presents an informal example, to identify the issues. Statistical models are often used to make causal inferences, for example, estimates of the impact of interventions: If we put a tariff of $10 a barrel on imported oil, how much will that affect the price level? the Gross National Product? If we spend another million dollars on schools, how much will that affect test scores?

Other kinds of causal inferences, more relevant to the stratification literature, do not feature such explicit interventions, but raise similar issues: How much of a salary difference between men and women is due to sex bias, and how much to differences in productivity? Where does ability count for more in getting high status jobs: Scotland or the United States?

In this section, I would like to present one highly stylized example, to focus the issues: How much does education affect income? Suppose we take a large random sample (so the distinction between parameters and estimates can be slurred over), and observe in the data that log income has a linear regression on education:

$$\log(\text{income}) = a + b \times (\text{education}) + \text{error}.$$

The errors seem to have nearly constant variance, and follow the normal curve quite nicely. We estimate $a$ and $b$ by ordinary least squares; $b$ turns out to be .05, and the conclusion is that sending people to school for another year will increase their incomes on average by 5%.

This conclusion, however, cannot rest on the way the data look, or even on replication across time or geography. It must depend on a theory of how the data came to be generated. (This is very well known to working social scientists, and I will sketch their argument in a moment; the novelty, if any, is only in the formulation.)

In effect, the theory has to be (or at least have as a consequence) that we are observing the results of a controlled experiment conducted by Nature. Subjects are given some number of years of schooling, and then the logs of their incomes are generated as if by the following two-step procedure:

(i) compute $a + b \times (\text{education})$;
(ii) add noise.

For want of a better term, I will refer to this procedure as a *linear statistical law,* connecting log income and education; and the whole thing can be called the *as-if-by-experiment* assumption. A slightly more dramatic theory: We might assume that Nature has randomly assigned people to the different educational levels—an *as-if-randomized* assumption. (Then the log-linear functional form can be estimated from the data.)

These sorts of theories seem to be implicit in the idea of "structural" or "causal" equations. Of course, it is impossible to tell just from data on the variables in it whether an equation is structural or merely an association. In the latter case, all we learn is that the conditional expectation of the response variable shows some connection to the explanatory variables, in the population being sampled. The decision as to whether an equation is structural must ride either on prior theory or on close examination of data outside the equation. Considering the impact of interventions is a useful armchair exercise to perform in trying to reach the decision, or at least figuring out what it means.

Coming back to income and education, it may be obvious by now that the equation is not structural, even if the data look just like textbook regression pictures. There is a third variable, family background, which drives both education and income. Our coefficient $b$ in the equation picks up the effect of the omitted variable, and is therefore a biased estimate of the impact of the proposed intervention—sending people back to school. Well, responds a strawman investigator, here's a path model to take care of the problem:

$$\text{education} = c + d \times (\text{family background}) + \text{error}$$
$$\log(\text{income}) = e + f \times (\text{education}) + g \times (\text{family background}) + \text{error}.$$

This model consists of two equations, assumed to be structural—as if the results of an experiment of nature. (These equations are hard to interpret on the as-if-randomized basis, but they do make sense—right or wrong—as linear statistical laws.) Anyway, says the strawman, now we've got it: $f$ represents the impact of education on income, with family background controlled for.

Unfortunately, it is not so easy. How do we know that it is right this time? What about age, sex, or ability, just for starters? How do we know when the equations will reliably predict the result of interventions—without doing the experiment? In a cartoon example, the objection is clear. In other cases, the point is easily lost in the shuffle. Sad experience shows that all too often, the real modelers pay as little attention to justifying their assumptions as the strawman, and their results are no more convincing.

Of course, the methodological issues are extremely perplexing, and any attempt to bring evidence to bear on the substantive issues deserves considerable sympathy. Duncan (1975) and Wright (1934) stressed the assumptions and the limitations of the technique. Simon (1954) presented a sophisticated defense of our strawman; and see Zellner (1984, chapter 1.4) for a critique of Simon. Many investigators seem to focus only on the statistical calculations. This essay can be viewed as an attempt to put the spotlight back on the assumptions.

## Regression Models

This section will discuss regression models when the explanatory variable $X$ is under experimental control, and modifications for observational data. Some threats to validity will be identified, and the first successful use of regression models will be noted.

To begin with, suppose that a variable $X$ is under experimental control, and it "causes" $Y$ in the sense that

$$Y = a + bX + U. \tag{1}$$

In this equation, $U$ is a random variable—like a draw made at random from a box of tickets. $Y$ is observable, but $U$ is not. The stochastic assumptions are as follows:

The distribution of $U$ is the same, no matter what value of $X$ is selected by the experimenter. (2.1)

The expected value of $U$ is 0. (2.2)

The variance of $U$ is finite. (2.3)

Each time the system is observed, an independent value of $U$ is generated. (2.4)

(The first and last of these are serious restrictions; the two middle ones have a more technical character.)

In this model, ordinary least squares (OLS) is a sensible way to estimate the parameters $a$ and $b$. (Optimality and robustness are among the least of our worries here.) Furthermore, the coefficient $b$ has a straightforward causal interpretation: It is the expected change in $Y$, if the experimenter intervenes and increases $X$ by one unit.

Of course, many interesting variables are not under experimental control. On the other hand, given the experimental model (1–2), it does not matter how the data on $X$ are collected, provided the distribution of $U$ is not disturbed. (An example to motivate the caveat: Selecting observational units on the basis of their $Y$-values can lead to systematic errors.)

Consider next a model like (1–2), but for observational data. This is a major transition; from now on, I will be discussing only studies where the investigator does not intervene to set the values of the explanatory variables. The first part of the model is a theory about the relationship between $X$ and $Y$, the two variables of interest. This can be stated, a bit quaintly, as follows: Nature selects $X$ according to some distribution, generates $Y$ according to the experimental model (1–2), and then presents the pair $(X, Y)$ to the investigator, but hides $U$. This is a linear statistical law connecting $Y$ and $X$.

Much (but not all) of this can be formalized starting from an equation like (1), with a slightly different interpretation:

$$Y = a + bX + U. \tag{3}$$

The equation says that the random variable $Y$ depends linearly on the random variable $X$, with an unobservable random error or disturbance term $U$. The parameters of this linear statistical law are $a$, $b$, and the variance of $U$. In this equation, $X$ is random.

The stochastic assumptions on $U$ and $X$ are parallel to the ones in the experimental model, except that (2.4) is dropped for now:

The disturbance term $U$ is independent of the explanatory variable $X$. (4.1)

The disturbance term $U$ has mean 0. (4.2)

The disturbance term and the explanatory variable have finite variance. (4.3)

The disturbance term $U$ is often interpreted as "the effect of omitted variables." If so, intervening to change $X$ should not change $U$, and this is perhaps the strongest form of the independence hypothesis (4.1). Also, the omitted variables must be assumed independent of the included variables. (See Pratt & Shlaifer, 1984 for discussion.)

With path models, one convention is to standardize the random variables $X$ and $Y$ in Equation (3) to have mean 0 and variance 1. The parameter $a$ will then be 0. (This convention will be followed here for expository reasons, and its logic will be reviewed later.) To make the variance of $Y$ come out to 1, another condition is needed:

$$\operatorname{var} U = 1 - b^2. \tag{4.4}$$

So far, $X$ and $Y$ are somewhat platonic. The next part of the model introduces data, an $X$-value and $Y$-value for each observational unit; these units will be indexed by $i$. This part of the model is intended to connect the data with the assumed relationship (3–4), and is the analog of (2.4). The data are modeled as observed values of pairs of random variables $(X_i, Y_i)$, which are independent from unit to unit, and obey the assumptions (3–4).

Technically, the conditions on the data are as follows:

The triplets $(X_i, U_i, Y_i)$ are independent across units $i$. (5.1)

For each $i$, the triplet $(X_i, U_i, Y_i)$ is distributed like $(X, U, Y)$ in (3–4). (5.2)

The explanatory variable is $X_i$, the response variable is $Y_i$, and the disturbance term $U_i$ is not observable. This sampling model is the basis for estimating $b$ from the data by OLS. (Of course, in the present setup it is the random variables that are standardized, not the data: Standardizing the data is one step in estimating the parameters. Also, some investigators prefer to condition on the $X$-values; this hardly affects the present argument, but would complicate the exposition, which attempts to define a sampling model and give theorems in terms of population parameters. Be the conditioning as it may, in the present setup the $X$-values are not under the investigator's experimental control, and that is what matters.)

To make the assumptions more vivid, imagine two boxes of tickets, an $X$-box and a $U$-box. In each box, the tickets have numbers on them. The tickets in the $X$-box average out to 0 with a variance of 1 (the standardization). The tickets in the $U$-box average out to zero with a variance of $1 - b^2$. The data are generated as follows. For each observational unit $i$, draw one ticket at random (with replacement) from the $X$-box and another, independently, from the $U$-box. The first ticket shows the value of $X_i$, and the second, $U_i$. Now use Equation (3) to compute $Y_i$ (Figure 1).

This set of assumptions lies behind the simple path diagram in Figure 2. The straight arrow leading from $X$ to $Y$ represents the $X$-term in Equation (3); the coefficient is shown near the arrow. The free arrow leading into $Y$ represents the disturbance term $U$.

The assumptions (3–4–5) are not explicit in the diagram. These assumptions are relatively strong, but something rather like them seems necessary to justify the full range of operations made by path analysts. For some applications, especially with relatively small samples, normality would have to be assumed too. (Of course, weaker assumptions can be used to justify partial conclusions in specific cases.)

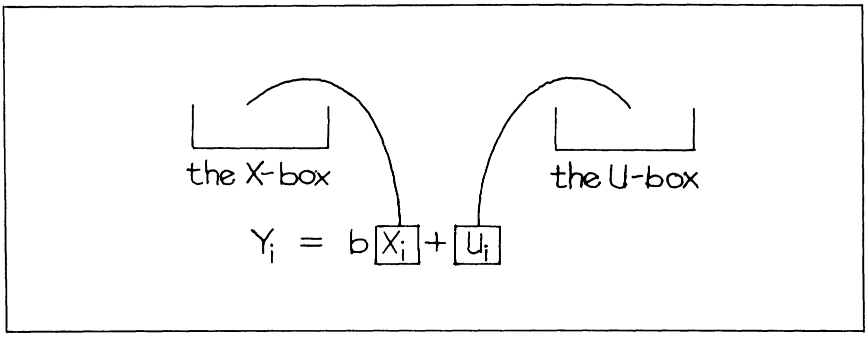



FIGURE 1. *The observational model*

The division of the model into two parts

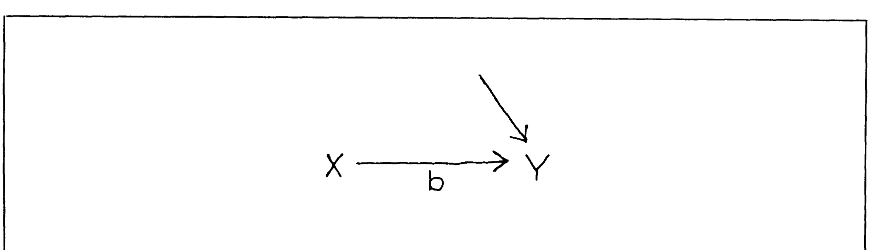
(i) the theoretical relationship between $X$ and $Y$,
(ii) the connection with the data on the observational units

is not standard but can be helpful analytically. For example, (3–4) may hold, but if the data are collected on a stratified sample then (5) fails. Or the data could be for a simple random sample so (5) holds, but the theoretical relationship (3–4) could be wrong.

Also, the two parts of the model play different roles. Part (i) underlies the causal interpretation of the parameter $b$, as the expected change in $Y$ if we intervene and change $X$ by one unit. The idea behind (i) is that changing $X$ by intervention makes no systematic impact on $U$, because the two are independent. So the expected change in $Y$ is just $b$ times the proposed change in $X$. On the other hand, part (ii) enables us to learn $b$ from the data by OLS.

In particular, a causal inference can be made from observational data—but its validity depends on the validity of the assumptions on the relationship between $X$ and $Y$ for the interpretation of $b$, and the connection with the data for its estimation by OLS. Together, the conditions say that

FIGURE 2. *A first path diagram*



the observational data on $X$ and $Y$ are generated as if by experiment, with Nature setting the values of $X$ by drawing them at random from a distribution, and then generating the $Y$s from the $X$s through a linear statistical law.

This set of conditions is somewhat more restrictive than assuming, for example, that $X$ and $Y$ are jointly normal and $(X_i, Y_i)$ are independent replicates of $(X, Y)$; and is not fully captured by (3–4–5). However, the as-if-by-experiment condition seems to be the hallmark of a structural equation or causal model, as opposed to a mere regression equation. One useful way to think about this distinction is to consider what the equation says about interventions.

The technical assumptions in a path model involve the chance behavior of the disturbance terms $U_i$. These terms are in principle not observable, so the assumptions are difficult to verify directly. Nor is the chance behavior of an unobservable quantity a topic that fires the imagination. Perhaps as a result, it is hard to get path analysts to focus on the chance assumptions. On the other hand, these assumptions do have empirical content—and should be tested before the model is taken seriously.

What are the main threats to the validity of the assumptions? Three will be mentioned now (cf. Tukey, 1954, p. 46):

(i) measurement error in $X$,
(ii) nonlinearity,
(iii) omitted variables.

Problem (i) is often recognized by workers in the field, and handled by "latent variable" models of the type popularized by Joreskog and Sorbom (1981) or Wold (1985); see Noonan and Wold (1983) for an example. However, the purported solution involves the introduction of yet another layer of assumptions, that there are repeated measurements linearly related to the latent variables. In my view, this only begs the question, by moving the kind of difficulties under discussion here to other (even less accessible) realms. See Freedman (1985) for a discussion.

Problems (ii) and (iii) will be discussed a bit abstractly in this section and the next, and illustrated on Hope's model. Take nonlinearity first. Instead of (1), suppose the generating equation is

$$Y = a + bX + cX^2 + U. \tag{6}$$

An investigator who fits a linear model like (1) will get an error term uncorrelated with $X$, but dependent on it. Then, changing $X$ must change the error in a systematic way, and the causal inference is invalid.

For an extreme case, suppose $X$ is uniformly distributed on the interval $[-1, 1]$, and $Y = X^2$. Fitting (1) gives $a = 1/3$ and $b = 0$, suggesting that a change in $X$ will cause no change in $Y$. That is clearly wrong: The effect of $X$ and $Y$ is all in the nonlinear error $U = X^2 - 1/3$. In particular, the path-analytic notion of "cause" is intimately bound up with linear statistical laws.

In this example, $b$ can be interpreted as the average change in $Y$ per unit change in $X$, that is, the average over $X$ of $d/dx\, E\{Y|X=x\}$. On this interpretation, however, $b$ depends on the distribution of $X$. And the whole "average change" idea breaks down if, for example, $U = X^3 - 3X/5$ and $Y = U$ so $a = b = 0$. The reason is that $U$ takes different values when $X = 1$ and $X = -1$ (cf. Tukey, 1954, p. 42).

Moving on to (iii), suppose the generating equation is (1), but the expected value of $U$ increases linearly with $X$. Then OLS will produce a biased estimate of $b$. Put another way, causal inference fails because changing $X$ makes a systematic change in $U$. One source of such dependence is omitted variables—problem (iii).

This problem too is well known to workers in the field, and their solution is to expand the system by adding more variables. That is what path models are all about. In sum, if the main variables in the system can be identified, and their causal ordering, and the form of the regression functions, the models can be more or less easily adapted. But there is a major difficulty: Current social science theory cannot deliver that sort of specification with any degree of reliability, and current statistical theory needs this information to get started.

Indeed, the method of least squares was developed by Gauss (1809/1963) for use in situations where measurements are well defined, where functional forms are dictated by strong theory, and where predictions are routinely tested against observations. The story is worth retelling: Astronomers discovered the asteroid Ceres while making telescopic sweeps, but lost it when it got too close to the Sun. Finding Ceres became one of the major scientific problems of the day.

To solve this problem, Gauss derived the equations connecting the observations on Ceres to the parameters of the orbit, using Newtonian mechanics. He then linearized the equations of motion and estimated the parameters by least squares, making careful estimates of the errors due to the linearization and due to random variation in the data. Finally, he used the equations to predict the current position of Ceres—a prediction borne out by astronomical observation.

In this example, Gauss started from well-established theory that specified the relevant variables and the functional form of their relationship. Much careful work had already been done on the error structure of the astronomical measurements. Finally, the model was tested against reality. These characteristics differentiate the original application of least squares from the application in path models. In situations where theory and measurement are less well developed, simpler and more informal statistical techniques might be preferable.

## Path Models

This section will develop path models, as linked sets of structural equations, with the response variable from one equation being an explanatory variable in a second. The assumptions will be highlighted, and the possibility of testing discussed; threats to validity will be discussed. The development is parallel to the one in the previous section, and it is convenient to use the example of $X$ and $Y$ in Equation (1). Suppose that there are two additional variables in the system, $Z$ and $W$. Suppose that together, $Z$ and $W$ cause $X$; then, $Z$ and $X$ cause $Y$; and the causation is through linear statistical laws. This theory about the relationships of the variables can be expressed in a path diagram (Figure 3).

The straight arrow leading from, for example, $Z$ to $X$ indicates that $Z$ appears in the equation explaining $X$; the free arrow leading into $X$ stands for the disturbance term in that equation. The path diagram, then, represents two linear statistical laws:

$$X = aZ + bW + U \tag{7.1}$$

$$Y = cX + dZ + V. \tag{7.2}$$

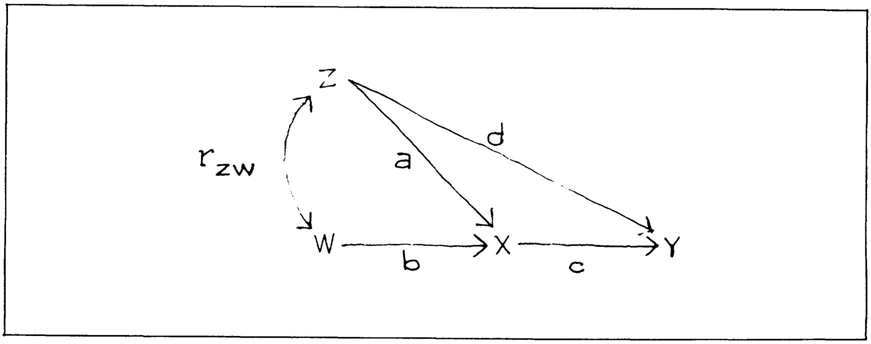
The random variables $Z$ and $W$ are "exogenous": They are viewed as causing the other variables in the model, but are not themselves explained. This is signaled by the curved, double-headed arrow in the diagram; next to the arrow is the correlation between these two variables. The variables $X$ and $Y$ are "endogenous"—explained within the model.

The random variables $U$ and $V$ in (7) are disturbance terms. Equation (7.1) relates $X$ to the exogenous variables; then (7.2) relates $Y$ to $X$ and the exogenous variables. As is usually said, the system is "recursive" rather than "simultaneous." (For more careful definitions, see the next section.)

Informally, Nature selects $(Z, W)$ from some distribution, and generates some noise $(U, V)$. The she or he computes $X$ and $Y$ from (7) and shows $(Z, W, X, \& \ Y)$ to the investigator. The disturbances $U$ and $V$ remain hidden.

The parameters $a, b, c, d$ in (7) are called "path coefficients" and are usually unknown. The following are the stochastic assumptions:

FIGURE 3. *A path diagram*



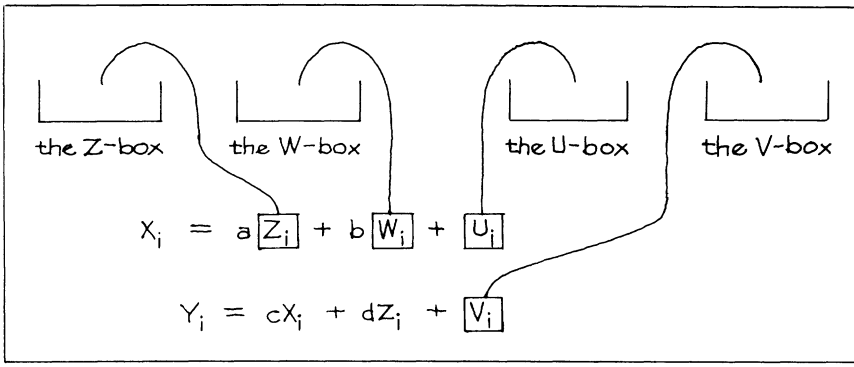



FIGURE 4. *Box model for path diagram*

The disturbance terms $U$ and $V$ are independent of each other and the exogenous variables $(Z, W)$. (8.1)

The disturbance terms $U$ and $V$ have mean 0. (8.2)

The disturbance terms and exogenous variables have finite variance. (8.3)

$Z$, $W$, $X$, and $Y$ are all standardized to have mean 0 and variance 1. (8.4)

Now for the connection between the theory and the data. There are $n$ observational units, indexed by $i$; for each unit, there are measurements on the four variables $Z$, $W$, $X$, and $Y$. These data are modeled as observed values of random variables $Z_i$, $W_i$, $X_i$, and $Y_i$, which are independent from unit to unit and obey the theory expressed in (7) and (8):

The six-tuplets $(Z_i, W_i, U_i, V_i, X_i, Y_i)$ are independent across units $i$. (9.1)

For each $i$, the six-tuplet $(Z_i, W_i, U_i, V_i, X_i, Y_i)$ is distributed like $(Z, W, U, V, X, Y)$ in (7–8). (9.2)

These assumptions are shown schematically in Figure 4.

For a moment, come back to the omitted-variables problem. If (7–8) are right, then a regression of $Y$ on $X$ alone gives a biased estimate of the effect of $X$, because the regression coefficient picks up part of the effect of the omitted variable $Z$. Specifying the right path model fixes this problem.

By virtue of assumption (9), the full system can be estimated by OLS. And the parameters of the linear statistical laws do have causal interpretations, as in the previous section. For example, suppose we intervene by keeping $W$ and $Z$ fixed but increasing $X$ by one unit: On average, this will cause $Y$ to increase by $c$ units—because the disturbance $V$ in $Y$ is unrelated to $Z$, $W$, or $X$, so the intervention has no systematic impact on

$V$. (The usual interpretation of the coefficients does present some difficulties, to be discussed below in the section on direct and indirect effects.)

Again, a causal inference has been made from observational data. This is, I think, the most attractive feature of the methodology. However, its validity depends on the assumption that the variables are related through linear statistical laws—and the path analyst got the variables and arrows right.

These assumptions are embodied in the path diagram and equations (7–8–9). They are inputs to the statistical analysis rather than outputs. All the statistical analysis can do (and it is no mean feat) is to estimate the correlations and path coefficients, or test that particular coefficients are zero—given the assumptions. Standard theory does not offer any strong tests of those assumptions; and the existing ones (plotting residuals, cross-validation, examining subgroups) are seldom done by path analysts.

The point may be a bit obscured by the somewhat technical way the term *causal inference* is being used. To restate matters: A theory of causality is assumed in the path diagram (the causal ordering, linear statistical laws, etc.). Within this context, what the path analysis does is to provide a quantitative estimate of the impact of interventions. The path analysis does not derive the causal theory from the data, or test any major part of it against the data. Assuming the wrong causal theory vitiates the statistical calculations.

## The Fundamental Theorem

One object of path analysis is to decompose correlation coefficients into additive components. Consider for now a general path diagram. The variables are still standardized, and the analogs of (7–8) are in force. The "fundamental theorem" is an identity among correlation coefficients and path coefficients.

First, some terminology and the assumptions. Say a variable $X$ is a "proximate cause" of $Y$ if there is a straight arrow leading from $X$ to $Y$ in the path diagram (i.e., $X$ appears in the regression equation explaining $Y$). By definition, an "exogenous" variable has no proximate cause whereas an "endogenous" variable has at least one proximate cause. (In this paper, I am taking the exogenous variables to be random, and will state the fundamental theorem and related decompositions at the level of population parameters, rather than sample estimates.)

There is a curved arrow joining every pair of exogenous variables, and one free arrow leading into each endogenous variable. But no free arrows lead into exogenous variables, or curved arrows into endogenous variables. Say $X$ is a "remote cause" of $Y$ if there is a sequence of straight arrows leading from $X$ to $Y$ in the diagram. Assume that the path diagram is recursive, in the following sense:

No variable can be a cause of itself, proximate or remote. (10)

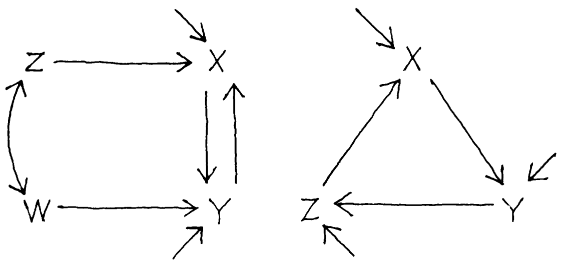



FIGURE 5. *Violating the conditions. In the left diagram* X *causes* Y *and* Y *causes* X. *At the right,* X *is a remote cause of itself.*

In particular, $X$ may be a remote cause of $Y$, or $Y$ a remote cause of $X$, or neither—but not both. Figure 5 shows two diagrams that violate the conditions.

(The left hand diagram in Figure 5 provides an example of "simultaneity": $X$ and $Y$ are obtained from $Z$ and $W$ by solving the pair of simultaneous linear equations:

$$X = aZ + bY + U$$

$$Y = cW + dX + V.$$

Such systems are common in econometric work, but less common in other social science fields. The new difficulty is that the right hand side variables become correlated with the errors, so OLS must be replaced by more complex estimation procedures.)

The "fundamental theorem" (Duncan, 1975, pp. 36ff or Wright, 1921) is the following:

Suppose $Y$ is endogenous and not a cause of $X$, proximate or remote. Then (11)

$$r_{XY} = \Sigma_Z p_{YZ} r_{XZ},$$

where

$p_{YZ}$ is the path coefficient from $Z$ to $Y$,
$r_{XY}$ is the correlation coefficient between $X$ and $Y$, and
the sum is over all $Z$ that are proximate causes of $Y$.

Technically, the set of exogenous variables is assumed independent of the disturbance terms, which are independent among themselves, as in (8.1).

It follows that an endogenous variable is independent of all the disturbance terms—except those among its causes. The identity (11) can be applied recursively, to express any correlation in terms of the path coefficients and the correlations among the exogenous variables. This leads to the deep "path tracing rule" of Wright (1921).

### Direct and Indirect Effects

One object of path analysis is to measure the "direct" and "indirect" effects of one variable on another. Take, for example, the path diagram in Figure 3. Successive applications of (11) show

$$r_{ZY} = d + ac + r_{ZW}bc.$$

Associated with this identity is some terminology:

$r_{ZY}$ is called the "total effect of $Z$ on $Y$,"
$d$ is the "direct effect of $Z$ on $Y$,"
$ac + r_{ZW}bc$ is the "indirect effect of $Z$ on $Y$," and
$ac$ is the "effect of $Z$ on $Y$ through $X$."

The relationship between the terminology and the diagram is pretty clear, but the connection to causal inference more problematic. Take the last effect first: We intervene by holding $W$ fixed but increasing $Z$ by one unit; this increases $X$ by $a$ units, which in turn makes $Y$ go up $ac$ units. So far, so good.

Now try the direct effect of $Z$ on $Y$: We intervene by fixing $W$ and $X$ but increasing $Z$ by one unit; this should increase $Y$ by $d$ units. However, this hypothetical intervention is self-contradictory, because fixing $W$ and increasing $Z$ causes an increase in $X$. Or is the disturbance term $U$ in (7.1) supposed to come down? How does that square with independence? Or the idea that $U$ represents "omitted variables?"

This may seem like a pointless tease about semantics, but given the research effort spent in composing and decomposing correlations, surely some attention to interpretation is called for. The only possibility for $d$ seems to be the rate of change of $E(Y \mid X, Z)$ with respect to $Z$, and calling this a "direct effect" is rather strange.

My view, stated in detail earlier, is that a path model represents the analysis of observational data as if it were the result of an experiment. At points such as this, it would be helpful to know more about the structure of such hypothetical experiments: What is to be held constant, and what manipulated?

### Standardizing the Variables

Should path models be given in terms of variables in their natural scale, or should they be standardized? The question is hardly a new one—see, for example, Achen (1982, p. 76), Blalock (1964, p. 51), Tukey (1954, p. 41), or Wright (1960). Apparently, standardizing only matters when comparing path coefficients estimated from different populations, that is, different distributions for the exogenous variables. And the answer depends on what is considered to be invariant across populations—that is, on the form of the social law assumed to govern the data. In other words, there is an empirical issue.

To focus ideas, consider the simple regression model of Equation (1). Let $X$ and $Y$ denote the variables in their natural scale (dollars, years of schooling completed, etc.); let $\tilde{X}$ and $\tilde{Y}$ denote the standardized variables. There are two versions of the equation, raw and standardized:

$$Y = a + bX + U \tag{1}$$

$$\tilde{Y} = \tilde{b}\tilde{X} + \tilde{U}. \tag{$\tilde{1}$}$$

Suppose first that equation (1) expresses a social law, with parameters $a$, $b$, and var $U$, which are invariant across populations ($X$-distributions). Now the path coefficient $\tilde{b}$ depends on the scale of $X$:

$$\tilde{b} = b\sqrt{\operatorname{var} X / [b^2 \operatorname{var} X + \operatorname{var} U]}.$$

Two investigators who work with different populations and standardize will get different path coefficients—and miss the invariance of $a$ and $b$. This is not a good research strategy.

On the other hand, the law governing the data could be in terms of standard units—Equation ($\tilde{1}$). That is, $\tilde{b}$ could be invariant across populations. Then fitting (1) in "natural" units will miss the invariance, because $a$ and $b$ vary across populations. Of course, the situation could be more complicated than (1) or ($\tilde{1}$).

Even if ($\tilde{1}$) is the right choice, the standard deviations of the disturbance terms in more complex models need not be invariant across populations. Consider, for example, the path model in Figure 3. Suppose that the law (7–8) underlying the social process is in standard units, so the variables $Z$, $W$, $X$, and $Y$ should be standardized. And the path coefficients will be stable across populations (joint distributions for $Z$ and $W$). But the standard deviations of the disturbance terms $U$ and $V$ depend on the path coefficients and $r_{ZW}$, so these standard deviations will change from population to population, because $r_{ZW}$ depends on the joint distribution of the exogenous variables.

The point can be illustrated on var $U$ in (7.1):

$$\begin{aligned} 1 &= \operatorname{var} X \\ &= a^2 \operatorname{var} Z + b^2 \operatorname{var} W + \operatorname{var} U \\ &\quad + 2ab \operatorname{cov}(Z, W) + 2a \operatorname{cov}(Z, U) + 2b \operatorname{cov}(W, U) \\ &= a^2 + b^2 + \operatorname{var} U + 2ab\, r_{ZW}. \end{aligned}$$

The first line holds by the standardization. For the same reason, in the second line, var $Z$ = var $W$ = 1 and cov $(Z, W) = r_{ZW}$. The other two covariances in line 2 vanish by (8.1). Now the equation can be solved for var $U$:

$$\operatorname{var} U = 1 - a^2 - b^2 - 2ab\, r_{ZW}.$$

Thus, the unexplained variation depends on the population being studied.

## Hope's Model

In this section, I will describe Hope's model, and then review it under the headings proposed earlier, stressing the weakness of the connection between the model and the underlying social process. One of Hope's technical innovations is the "autonomy coefficient," which may be viewed as an attempt to deal with certain kinds of measurement error. It will be discussed too. Along the way, I will try to give the flavor of the conclusions drawn from the model.

First, some background for the model. Since the publication of the Coleman et al. report (1966), there has been an extended debate concerning the impact of schools on the educational attainments of students, and the achievements afterward. Jencks et al. (1972) is often cited in this connection. Chapter 1 of Hope (1984) is a contribution to this literature. It addresses the following sorts of questions: Do schools matter? Does education have more of an impact in Scotland or in the United States? Of course, to answer these questions, the terms have to be defined, and background variables controlled.

Hope measures outcomes in terms of the occupations of his subjects. Only secondary education is considered. The background variables are two: IQ and father's occupation. The American data are drawn from Jencks and will not be discussed here. For Scotland, Hope measures occupations on a 9-point scale, ranging from "professionals and large employers" at the top to "agricultural workers" at the bottom (1984, p. 17). For the path model, occupations were ranked according to "social standing" by 12 college students, and the principal component of the 12 rankings was used as the occupation variable (p. 18).

Secondary education in Scotland was on a track system, and Hope measures this variable on a 7-point scale, according to the track taken by the subjects (1984, p. 14). The top track "completed five years of secondary education in a general course with two foreign languages." The bottom track "three years of secondary education in a modified class (for less able and backward children in an ordinary school)."

The data are from the "Scottish Mental Survey." The sample was drawn in 1947, and consists of all 11-year-old boys born on the first day of every other month. The children were followed until 1964, and data are available on nearly 600 of them, including anthropometry, scores on Form L of the Stanford-Binet IQ test, and a "sociological schedule" that included father's occupation. Sample attrition was less than 10%.

Hope proposes a path model for his four variables; it is shown in Figure 6, with coefficients estimated from the American and Scottish data. There are two structural equations:

$$\text{education} = a \times \text{IQ} + b \times (\text{father's occupation}) + \text{error} \tag{12.1}$$

$$\text{occupation} = c \times (\text{education}) + d \times \text{IQ} + e \times (\text{father's occupation}) + \text{error}. \tag{12.2}$$

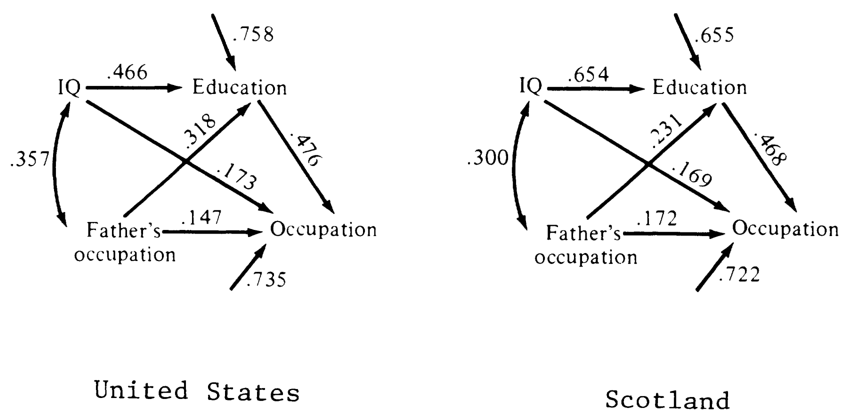
(In these equations, of course, IQ, education, and occupation are attributes of the sons.) This model will now be reviewed under the headings proposed earlier, starting with measurement issues.

The difficulties in measuring intelligence and success in life are only too well known. But what about education? For one thing, the school variable in the United States is quite different from the one in Scotland (years completed rather than track). And this kind of difference can have a substantial impact on the path coefficients. (A similar issue is noted in the section on standardization, above.)

Moreover, the school variable in Scotland includes not only characteristics of the schools, but also characteristics of the students. For example, the difference between tracks 1 and 3 lies in whether the students completed the course: This is a measure of students' ability and character, as well as of the education received. The path analysis is being used to separate the inputs to the school from outputs, but the two are already entangled in the school variable—before the path analysis can do anything.

This point has been questioned by some readers. To make the issue clearer, suppose the measured school variable is just another IQ score—like form M instead of form L. None of the analysis would then have anything to do with education. For a second example in a similar vein, suppose the path model in Figure 6 is right—for some variable that represents

FIGURE 6. *Hope's path model*




Source: Hope (1984, pp. 26–27).

educational quality. In this second hypothetical world, the path from IQ to education represents the fact that brighter students choose better schooling, on average. Suppose too that the measured school variable is a linear combination of IQ and the underlying educational quality variable. In this second hypothetical, the estimated path coefficients are substantially biased; and the bias depends on the unknown error structure in the measurements.

One of Hope's technical innovations might be viewed as an attempt to get around the second sort of measurement problem. He decomposes the direct effect of education on occupation into an "autonomous" and a "heteronomous" component (1984, pp. 9, 24ff). To define these terms, assume the relationships part of the model in Figure 6, that is, the analog of (7–8). Let $E$ and $O$ be the endogenous variables, education and occupation; let the exogenous variables be $I$ and $F$, IQ and father's occupation. Hope's decomposition can be expressed as an equation:

$$p_{OE} = p_{OE}\,v + p_{OE}\,p_{EI}\,r_{EI} + p_{OE}\,p_{EF}\,r_{EF}. \tag{13}$$

The first term on the right is the "autonomous effect" of education on occupation; $v$ is the variance of the disturbance term in the equation for $E$.

To understand this term geometrically, let $\hat{E}$ and $\hat{O}$ be the projections of $E$ and $O$ on the space spanned by $I$ and $F$; let $E^{\perp} = E - \hat{E}$ and $O^{\perp} = O - \hat{O}$, so $E^{\perp}$ and $O^{\perp}$ are the projections on the orthocomplement. In other words, $E^{\perp}$ and $O^{\perp}$ represent education and occupation, net of IQ and father's occupation: "net" means, after subtracting out the projections. The "autonomous effect" of education on occupation, that is, the first term on the right of (13), turns out to be the covariance of $O^{\perp}$ and $E^{\perp}$. Since $E^{\perp}$ is not standardized, the covariance is not a regression coefficient; apparently, division by $v = \operatorname{var} E^{\perp}$ just gives you back $p_{OE}$, as is also clear from (13).

(The remaining two terms in the equation seem to be part of the "heteronomous effect"; however, I see no direct geometric interpretation for them, nor do Hope's verbal descriptions in his table 1.6 make much sense to me, let alone the identification of these terms with the "universalistic" and "particularistic" in education.)

In principle, it is harmless to compute a covariance. But what is Hope's interpretation?

> The aim of the research is . . . to quantify the effects of education over and above any effects which it transmits from input variables [$I$ and $F$]. To accomplish that aim, we must ask ourselves a very simple question: To what extent is education acting as a transmitter (heteronomously) and to what extent is it contributing its own autonomous effect, over and above the transmitted effects, to the process of stratification? (1984, p. 9)

There is at least one peculiarity in this interpretation: Suppose the path model in Figure 6 is right. If we intervene by holding IQ and father's occupation fixed but increase education, then it is $p_{OE}$ that measures the change in the subject's occupation, not the "autonomous effect." Similarly, if the measured school variable is a linear combination of IQ and the underlying educational quality variable, the autonomy coefficient is quite a biased estimate of the path coefficient for educational quality. On the other hand, if the measured school variable is just another IQ score, the autonomy coefficient may say something about the relationship of IQ tests to occupational achievement—but nothing about schools, because these are absent from the model. Statistical computations are no substitute for a proper specification.

Hope cites Finney (1972) in defense of the autonomy coefficient, but Finney's message is more that the "indirect effect" of an exogenous variable on an endogenous one includes the correlations among the exogenous variables, and is therefore noncausal; also, this indirect effect depends on the population, that is, the joint distribution of the exogenous variables, and so is not comparable across populations (cf. previous sections on direct effects and standardization). Finney's second criticism applies directly to Hope's autonomy coefficient.

Now let me quote Hope (1984) again:

> The basic idea [of the autonomy coefficient] can be quite adequately represented by the following simple analogy. If we think of the paths in a path model as pipes along which water flows, and if we imagine one pipe connecting A and B, and another connecting B and C, then we naturally wonder whether the flow between B and C is entirely accounted for by the flow between A and B, or whether more water comes into the system at B. Such incoming water models our idea of the autonomous effect of B which, in our application of the analogy, stands for education. (p. 8)

This analogy does not really serve to differentiate among disturbance terms, direct effects, or autonomy coefficients. Indeed, since correlation and regression coefficients describe changes rather than levels, the passage seems to be a confusion: The statistical constructs in the path model relate less to the flow of water than to variations in the flow—ripples.

This completes the discussion of measurement issues and the autonomy coefficient. Now we come to the stochastic assumptions. Why do the variables satisfy a linear statistical law? Nowhere does Hope ask this question. With a 7-point scale for education, linearity is hard to take seriously; and Hope himself points to noticeable skewness in the IQ data. Also, the impact of the fast track may be larger on the brighter students. Heteroscedasticity is another problem: American data suggest that variance of intelligence and schooling depend on occupational level (see, e.g., Crouse & Olneck, 1979).

Omitted variables must be considered too. Is the process of social stratification the same in the Highlands as in the cities? Geography does not appear in the model. Are schools in Edinburgh similar to those in Glasgow? Schools per se do not appear in the model, except through the tracks variable. To have schools omitted is peculiarly ironic in a study of their effects, especially when the author thinks (as Hope does, see pp. 19–20) that it is the individual characteristics of different schools that make for strong educational effects.

Given such problems, what connects the model to reality? What makes Hope think the assumptions hold? I could only find a few sentences responsive to these questions, and quote them in full (1984, pp. 14–15):

> The theoretical model that informs our analysis of the effects of education on stratification was implicit in the design of the Scottish Mental Survey (Hope, 1980). It postulates that, for boys, the significant inputs they bring to the social system are cleverness, character, and class. Since we lack comparative data on character, we omit that variable from the current model (it was included in the model in Hope, 1980).

At best, this passage justifies including IQ and father's occupation in (12). It does not justify the linearity or the stochastic assumptions. Nor does it justify treating the equations in (12) as structural, rather than mere associations; on the contrary, it flags another important missing variable. And so in the end Hope does not connect Figure 6 with real boys who go to school, graduate, and get jobs. Neither does the cited article (Hope, 1980), which presents a path model rather like Figure 6, but including variables labeled "personality" and "qualifications." In that article, there is much ingenuity devoted to the geometry of factor analysis—and none to elucidating the relationship between the geometry and the boys.

Hope is not alone in these respects. Indeed, I do not think there is any reliable methodology in place for identifying the crucial variables in social systems or discovering the functional form of their relationships. In such circumstances, fitting path models is a peculiar research activity: The computer can pass some planes through the data, but cannot bind the arithmetic to the world outside.

At such points as this, modelers will often explain that nothing is perfect, all models are approximations, so maybe Hope's model is good enough: How much difference can the blemishes make? In this particular case, blemishes could really matter, because the differences in path coefficients between Scotland and the U.S. are small (the largest is for the direct effect of IQ on education). And substantial conclusions are drawn from these differences:

> What we have shown is that Scotland . . . is more merit[ocratic] than the United States. Taking transmission of IQ as universalistic and transmission of father's occupation as particularistic, we may say that the ratio of universalism to particularism in education . . . is 2 to 1 in the United States and more than 4 to 1 in Scotland. The overall direct effect of education on occupation is about .47 in both. Within this total, the ratio of autonomous effect : universalism : particularism is 57 : 27 : 15 in the United States, as against 43 : 47 : 10 in Scotland. We conclude, therefore, that data which have previously been held to manifest negligible autonomous effects of schools (in the United States) in fact ascribe a stronger effect to schools

> than they do to the characteristics of students entering those schools. (Hope, 1984, p. 30)

This sort of finding has to be really sensitive to the rather arbitrary specification. To illustrate the sensitivity: Adding "personality" and "qualifications" to the model, as in Hope (1980), makes the direct effect of education on occupation drop from .47 to .23. (This is rather close to the autonomy coefficient, but I do not see much logical connection between adding those two variables and projecting out the two exogenous variables.)

The differences in path coefficients between Scotland and the United States may be largely due to the scaling, or the differential impact of omitted variables and measurement error. In any case, the central issue is what connects the path model and the process by which boys get jobs; this connection is simply not established. Assumptions matter, and with path models it is too easy to lose track of this.

Modelers may then explain that they are only doing data reduction. Let us agree for the moment that Hope's data look more or less like a sample from the multivariate gaussian distribution, and that standardization is appropriate. (Otherwise, his analysis goes off the rails almost immediately.) Now the data for each country can be summarized by a table of six correlation coefficients. However, the path model has six parameters too, and Hope's table 1.6 analyzes these into a dozen components. Data reduction is not the game here.

Other readers may feel that nobody could be taking the models so seriously; after all, a model is just some way of looking at the data. Well, here is Hope's view:

> We begin our study by asking whether Scotland was indeed the meritocracy it is often alleged to have been. In the course of answering this question we will refine the definition of meritocracy to the point where its presence or absence, or rather the degree to which it is present, can be assessed in precise, quantitative terms. Of course, no such quantification can be final or irrevocable; nevertheless, it has distinct advantages over imprecise and impressionistic statements. In the first place, it gives us some idea of an order of magnitude we did not possess before. Second it enables us to compare degrees of magnitude in different societies. And in the third place it is disputable on empirical grounds and corrigible according to rational criteria of evidence and rebuttal. But the really significant effect of quantification is, or ought to be, none of these. Rather does it lie in the effort at refining and exploring the meaning of analytical concepts which employment of a model calls for. . . . But in so constraining them we will make every effort to see to it that meaning is not warped beyond the bounds of normal usage, but rather is tightened up in a way which will command general approval. (1984, p. 6)

Hope is doing something much more ambitious than data analysis. But his statistical technique has led him astray, and he almost knows it:

> It will become apparent to the reader that the following work is not modest in its aims. And to those who observe that extensive conclusions are built on fairly exiguous foundations, the author can only reply that this is indeed the case. (1984, p. 3)

Hope's starting point may have been that a son's intelligence and his father's occupation jointly influence the boy's education, and then all three factors influence the boy's choice of occupation—influence but do not determine. Look back at Figure 6. Fixing IQ, fathers' occupational status, and education still leaves about 70% of the variation in sons' occupational status, both in the U.S. and in Scotland—in good agreement with Blau and Duncan (1967, p. 170).

These seem to me to be quite interesting facts, not much affected by the difficulties in path modeling under discussion. Scotland may be more meritocratic than the U.S., but on the evidence of Figure 6 neither country is exactly caste-ridden; rough insights whose value is considerable. (Something like this is at the core of Blau & Duncan, 1967.)

There are now some interesting questions, and to answer them, Hope takes for granted that his measurements on the variables are connected through linear statistical laws. He can hardly be blamed for doing so, because nearly everyone does the same; but it is precisely the move from rough insight to full-blown path model that seems so counterproductive to me.

The path model might be useful to predict the results of interventions, or of changing circumstances, or to provide a better understanding of the stratification process. But this sort of interpretation must ride on a theory, because something is needed to connect the statistical calculations to the process. As far as I can see, this theory is exactly what is missing.

## Meta-Arguments

For a historical discussion of path models in sociology, see Bernert (1983). These models were developed by Wright (e.g., 1921, 1934) for use in genetics. But later applications even in that field remain controversial: see Karlin (1979), or Karlin, Cameron, and Chakraborty (1983), with discussion by Wright et al. Tukey (1954, pp. 60–66) found the method attractive, but had doubts about the one specific example he presented.

Many investigators have written about the problem of drawing causal inferences from observational data. Some have stressed the as-if-randomized assumption: for example, see Holland (1986) or Pratt and Shlaifer (1984). Others have focused on the weaknesses of causal models: see Baumrind (1983), Cliff (1983), de Leeuw (1985), and Ling's (1983) scathing review of Kenny (1979). Lieberson (1985) is quite skeptical about the possibility of making statistical adjustments that bring observational data into the as-if-randomized condition.

Econometrics is an interesting test case for the modeling approach, because the technique is extremely sophisticated, and commercial services using macro-models make real forecasts whose accuracy can be tracked: see Christ (1975), Litterman (1986), McNees (1979, 1986), and Zarnowitz (1979). Basically, the major forecasting models do not do at all well unless their equations are revised frequently, and the intercepts reestimated subjectively by the modelers. Even then, such modeling groups do no better than the forecasters who proceed without models. Finally, the different modeling groups tend to make quite similar forecasts; but their models often disagree sharply about the projected impact of policy actions.

There are prominent critiques of standard econometrics and the underlying data by insiders, starting with the classic exchange between Keynes (1939, 1940) and Tinbergen (1940). More recent citations are Hausman and Wise (1985), Hendry (1980), Leamer (1983), Leontief (1971), Lucas and Sargent (1979), Morgenstern (1963), and Sims (1980). In some cases, of course, the proposed cure may be worse than the disease.

My own views have been argued in Daggett and Freedman (1985), and Freedman (1985), with discussion by Joreskog and Fienberg; in the context of energy models, see Freedman (1981), or Freedman, Rothenberg, and Sutch (1983), with discussion by Hogan and Smith; in the context of census adjustment, see Freedman and Navidi (1986), with discussion by Kadane et al.

One line of defense against foundational attack is Bayesian, as in Sims (1982). The disturbance terms are held to represent not omitted variables but a component of subjective uncertainty; another component of subjective uncertainty is expressed by a prior on the parameters of the model. The functional form of the model is common ground, on which all such Bayesian econometricians meet. Given the great diversity of functional forms in the econometric literature, and their transience, his argument does not seem to reach an important question: What is the relevance of textbook linear models to the economy?

Another interesting line of defense is sketched by Achen (1982), who says that we can use data to learn about social process. He proves the point, drawing on Veblen (1975) to make a charming and prima facie persuasive argument that the Manchester *Union-Leader* influenced elections in New Hampshire. A major tool is regression equations. He and Veblen think through a variety of qualitative positions about the political process, and their implications for the regressions. Then they look at the data, and only one position survives the test. But as Achen notes (p. 29), there is no pretense of developing a structural model; the equations are purely descriptive.

In some circumstances, regression equations are useful ways to look at data; the coefficients and the standard deviation around the regression plane can be good summary statistics. This is so, at least when the data look something like a sample from a multivariate gaussian distribution. Achen and Veblen have one success story along these lines. So did Blau and Duncan (1967). And so does Hope, as noted above. But there is a real difference between summarizing data and building path models. To take a simpler example, you can report a batting average without adopting countably additive probability theory and the strong law of large numbers.

The critique of path models in this paper is hardly original. A few words more about some of the conventional responses may be in order. Models are often held to be of value in offering guidelines to decisionmakers rather than descriptions or forecasts of behavior. The test often proposed for such models is whether decisionmakers find them useful. (The argument is strongest for normative models, but is often made for descriptive models as well.) In the present context, this argument seems circular.

Decisionmakers may like path analysis because the models appear to provide answers to important questions, but it is the use of the models by academic investigators that makes the process respectable. (For example, there are plenty of chartists on Wall Street, but their technique is not prominent in our literature.) The real issue is old-fashioned: Why are the answers from the path model at all dependable? In the end, we still need to verify the model's assumptions, by some combination of theory and practice, before recommending its conclusions to the decisionmaker.

We arrive, then, at the descriptive aspect. Here, two points are often made:

(i) Valid inferences can be drawn from axioms known to be false.

(ii) No model is perfect; what counts is that the model should be better than the next-best alternative.

Both points are quite weak. In the present context, the first is almost irrelevant, because the conclusions of path models are so close to their assumptions. More generally, argument (i) can only shift the burden from the assumptions to the conclusions. After all, if the modeler doesn't believe those assumptions, they can hardly be part of his or her reason for believing the conclusions.

Even in the best of theories, individual axioms may not be testable, so the modeler has to consider several axioms in combination and derive testable implications from the set: compare Blalock (1969, chapter 2). So there is a grain of truth to argument (i). However, the street version almost relies on a parody: If $A$ then $B$, and $A$ is false, therefore $B$. The anti-syllogism can be traced back to Friedman (1953), but that time he had to be kidding us.

Argument (ii) has some merit in some situations, but its relevance to statistical modeling in the social sciences is slight. The reason is that there is a wide range of imperfection in human knowledge, and it is not so obvious where to locate path models in that spectrum. The results of a path analysis depend for their validity on some underlying causal theory. If the theory is rejected, the interpretations have no foundation. Why, then, should they be held to dominate alternatives? And other modes of enquiry do exist: De Tocqueville was making comparative studies of the new world and the old long before path models arrived on the scene.

## Conclusion

This kind of negative article may seem incomplete. Path analysts will ask, not unreasonably, "Well, what would you do?" To this question, I have no general answer, any more than I can say in general how to do good mathematical research. Still, social science *is* possible, and needs a strong empirical component. Even statistical technique may prove useful—from time to time.

There are some such techniques on the horizon, which do not depend on prior specification the way path models do: see Asimov (1985), Breiman, Friedman, Olshen, and Stone (1984), Fisherkeller, Friedman, and Tukey (1975), Huber (1985), or McDonald (1984). It remains to be seen whether these innovations can be used to make improvements over path analysis. They do offer better data-analytic capabilities. Log linear models should also be mentioned (and their problems noted).

My opinion is that investigators need to think more about the underlying social processes, and look more closely at the data, without the distorting prism of conventional (and largely irrelevant) stochastic models. Estimating nonexistent parameters cannot be very fruitful. And it must be equally a waste of time to test theories on the basis of statistical hypotheses that are rooted neither in prior theory nor in fact, even if the algorithms are recited in every statistics text without caveat.

## References

Achen, C. (1982). *Interpreting and using regression.* Beverly Hills, CA: Sage.

Adams, R., Smelser, N., & Treiman, D. (1982). *Behavior and social science research: A national resource.* Washington, DC: National Academy of Sciences Press.

Asimov, D. (1985). The grand tour. *SIAM Journal of Scientific and Statistical Computing, 6,* 128–143.

Baumrind, D. (1983). Specious causal attribution in the social sciences: The reformulated stepping-stone theory of heroin use as exemplar. *Journal of Personality and Social Psychology, 45,* 1289–1298.

Bernert, C. (1983). The career of causal analysis in American sociology. *The British Journal of Sociology, 34,* 230–254.

Blalock, H. M. (1964). *Causal inferences in non-experimental research.* Chapel Hill: University of North Carolina Press.

Blalock, H. M. (1969). *Theory construction.* Englewood Cliffs, NJ: Prentice Hall.

Blau, P. M., & Duncan, O. D. (1967). *The American occupational structure.* New York: Wiley.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees.* Belmont, CA: Wadsworth.

Christ, C. (1975). Judging the performance of econometric models of the United States economy. *International Economic Review, 16,* 54–74.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research, 18,* 115–126.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Wood, A. M., Weinfield, F. D., & York, R. L. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Department of Health, Education, and Welfare.

Crouse, J., & Olneck, M. (1979). The IQ meritocracy reconsidered. *American Journal of Education, 88,* 1–31.

Daggett, R., & Freedman, D. (1985). Econometrics and the law: A case study in the proof of antitrust damages. In L. LeCam & R. Olshen (Eds.), *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer* (Vol. I, pp. 126–175). Belmont, CA: Wadsworth.

de Leeuw, J. (1985). Review of books by Long, Everitt, Saris and Stronkhorst. *Psychometrika, 50,* 371–375.

Duncan, O. D. (1975). *Introduction to structural equation models.* New York: Academic Press.

Finney, J. M. (1972). Indirect effects in path analysis. *Sociological Methods and Research, 1,* 176–186.

Fisherkeller, M. A., Friedman, J. H., & Tukey, J. W. (1975). *PRIM-9: An interactive multidimensional data display and analysis system* (Report SLAC-PUB-1408). Stanford, CA: Stanford Linear Accelerator Center.

Freedman, D. (1981). Some pitfalls in large econometric models: A case study. *Journal of Business, 54,* 479–500.

Freedman, D. (1985). Statistics and the scientific method. In W. Mason & S. Fienberg (Eds.), *Cohort analysis in social research: Beyond the identification problem* (pp. 345–390). New York: Springer.

Freedman, D., & Navidi, W. (1986). Regression models for adjusting the 1980 census. *Statistical Science, 1,* 3–39.

Freedman, D., Rothenberg, T., & Sutch, R. (1983). On energy policy models. *Journal of Business and Economic Statistics, 1,* 24–36.

Friedman, M. (1953). *Essays in positive economics.* Chicago: University of Chicago Press.

Gauss, C. (1963). *Theoria motus corporum coelestium.* New York: Dover. (Original work published 1809.)

Hausman, J., & Wise, D. (1985). Technical problems in social experimentation: Cost versus ease of analysis. In J. Hausman & D. Wise (Eds.), *Social experimentation.* Chicago: University of Chicago Press.

Hendry, D. (1980). Econometrics—Alchemy or science? *Econometrica, 7,* 387–406.

Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81,* 945–970.

Hope, K. (1980). A geometrical approach to sociological analysis. *Quality and Quantity, 14,* 309–325.

Hope, K. (1984). *As others see us: Schooling and social mobility in Scotland and the United States.* New York: Cambridge University Press.

Huber, P. (1985). Projection pursuit. *Annals of Statistics, 13,* 435–475.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America.* New York: Basic Books.

Joreskog, K. G., & Sorbom, D. (1981). *LISREL—Analysis of linear structural relationships by the method of maximum likelihood.* Chicago: International Educational Services.

Karlin, S. (1979). Comments on statistical methodology in medical genetics. In C.

F. Sing & M. Skolnick (Eds.), *Genetic analysis of common diseases: Application to predictive factors in coronary disease* (pp. 497–520). New York: Liss.

Karlin, S., Cameron, E. C., & Chakraborty, R. (1983). Path analysis in genetic epidemiology: A critique. *American Journal of Human Genetics, 35,* 695–732.

Kenny, D. (1979). *Correlation and causation.* New York: Wiley.

Keynes, J. (1939). Professor Tinbergen's method. *The Economic Journal, 49,* 558–570.

Keynes, J. (1940). [Comment on Tinbergen's response.] *The Economic Journal, 50,* 154–156.

Leamer, E. (1983). Taking the con out of econometrics. *American Economic Review, 73,* 31–43.

Leontieff, W. (1971). Theoretical assumptions and nonobserved facts. *American Economic Review, 61,* 1–7.

Lieberson, S. (1985). *Making it count.* Berkeley: University of California Press.

Ling, R. (1983). Review of correlation and causation by Kenny. *Journal of the American Statistical Association, 77,* 489–491.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions—Five years of experience. *Journal of Business and Economic Statistics, 4,* 25–38.

Lucas, R. E., & Sargent, T. J. (1979). After Keynesian macroeconomics. *Federal Reserve Bank of Minneapolis Quarterly Review,* 3.

McDonald, J. A. (1984). *Interactive graphics for data analysis* (Report ORION 11). Stanford, CA: Stanford University, Dept. of Statistics.

McNees, S. (1979). The forecasting record for the 1970's. *New England Economic Review.*

McNees, S. (1986). Forecasting accuracy of alternative techniques: A comparison of U.S. macroeconomic forecasts. *Journal of Business and Economic Statistics, 4,* 5–24.

Morgenstern, O. (1963). *On the accuracy of economic observations* (2nd ed.). Princeton, NJ: Princeton University Press.

Noonan, R., & Wold, H. (1983). Evaluating school systems using partial least squares. *Evaluation in Education, 7,* 219–364.

Pratt, J. W., & Shlaifer, R. (1984). On the nature and discovery of structure. *Journal of the American Statistical Association, 79,* 9–21.

Simon, H. (1954). Spurious correlation: A causal interpretation. *Journal of the American Statistical Association, 49,* 467–479.

Sims, C. (1980). Macroeconomics and reality. *Econometrica, 48,* 1–49.

Sims, C. (1982). *Scientific standards in econometric modeling.* Paper presented at the symposium on the Developments in Econometrics and Related Fields, Netherlands Econometrics Institute.

Tinbergen, J. (1940). [Reply to Keynes.] *The Economic Journal, 50,* 141–154.

Tukey, J. W. (1954). Causation, regression and path analysis. In O. Kempthorne, T. A. Bancroft, J. W. Gowen, & J. L. Lush (Eds.), *Statistics and mathematics in biology* (pp. 35–66). Ames: Iowa State College Press.

Veblen, E. (1975). *The Manchester Union-Leader in New Hampshire elections.* Hanover, NH: University Press of New England.

Wold, H. (1985). Partial least squares. In J. Kotz & N. L. Johnson (Eds.), *Encyclopedia of social sciences* Vol. 6, pp. 581–591. New York: Wiley.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research, 20,* 557–585.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5,* 161–215.
Wright, S. (1960). Path coefficients and path regressions. *Biometrics, 16,* 189–202.
Zarnowitz, V. (1979). An analysis of annual and multiperiod quarterly forecasts of aggregate income, output, and the price level. *Journal of Business, 52,* 1–34.
Zellner, A. (1984). *Basic issues in econometrics.* Chicago: University of Chicago Press.

## Author

D. A. FREEDMAN, Professor of Statistics, University of California, Berkeley, CA 94720. *Specializations:* statistics: modeling, foundations.