Structural Equations and Path Analysis for Discrete Data[1]

Christopher Winship
Northwestern University and *National Opinion Research Center*

Robert D. Mare
University of Wisconsin—Madison

This article proposes a solution to the long-standing methodological problem of incorporating discrete variables into causal models of social phenomena. Only a subset of the variety of ways in which discrete data arise in empirical social research can be satisfactorily modeled by conventional log-linear or logit approaches. Drawing on the insights of several literatures, this article exposits a general approach to causal models in which some or all variables are discretely measured and shows that path analytic methods are available which permit quantification of causal relationships among variables with the same flexibility and power of interpretation as is feasible in models that include only continuous variables. It presents methods of identifying and estimating these models and shows how the direct and indirect effects of independent variables can be calculated by extensions of usual path analysis methods for continuous variables. An important distinction developed here is that discrete variables can play two roles: (1) as measures of inherently discrete phenomena and (2) as indicators of underlying continuous variables. The value of this distinction is shown in two empirical examples examined previously by other authors. In examining the effects of social background and parental encouragement on college plans of high school seniors, the article shows that modeling a discrete measure of encouragement as an indicator of a latent continuous variable rather than as an inherently discrete variable (as has been done in previous analyses) provides a clearer interpretation and a superior fit to the data. In examining the effects of state Fair-Employment-Practices Legislation on black-white wage differentials, this study shows that two distinct effects on the relative wage can be detected: the direct ameliorative effect of the law itself and the effect of the popular progressive sentiment for racial equality of which the law is an indicator. The methods and models presented here are not only natural generalizations of structural equation and path analysis methods for continuous variables to include discrete variables but also provide a means of investigating a richer variety of substantive hypotheses than is feasible with methods for discrete data commonly used in the sociological literature to date.

[1] An earlier version of this paper, entitled "Structural Equation Models for Discrete Data," was presented at the meetings of the American Sociological Association at Toronto, Canada, August 25, 1981. This research was supported by National Science Foundation grant

0002-9602/84/8901-0002$01.50

One of the most important methodological developments in the social sciences in recent decades has been the use of structural equation models to specify and interpret causal relationships among variables (Goldberger and Duncan 1973; Duncan 1975; Bielby and Hauser 1977). Following upon related developments in other disciplines (Goldberger 1971), structural equation models in sociology have been applied primarily to processes in which the dependent variables are continuously scaled. Thus these models have been amenable to empirical analyses that use the general linear model and its extensions. More recently, causal analysis has been applied to discrete response variables (Goodman 1972, 1973*a,* 1973*b,* 1979; Fienberg 1980). Goodman and others have shown how to specify causal relationships among discrete variables and quantify the relationships by application of log-linear or logit models. These methods permit analysts to impose causal interpretations on multivariate systems of discrete variables; illuminating applications now appear in the literature (e.g., Gortmaker 1979).

At the same time, however, commentators have pointed out that causal models constructed from log-linear and logit models have limitations and that these models are not directly analogous to causal models with continuous variables (Fienberg 1975, 1980; Rosenthal 1980). Although there has been considerable interest in developing both a path analysis for discrete variables and one based on log-linear and logit models (e.g., Leik [1975], and the series of articles in *Sociological Methodology 1976* on this subject [Heise 1975]), no one has yet presented a satisfactory solution to this problem.[2] Indeed, several commentators have suggested that such a path analysis cannot be developed (Davis and Schooler 1974; Davis 1975; Fienberg 1980).

The limitations of the log-linear and logit models derive in part from the nonlinear functional forms of relationships among discrete variables that are implicit in systems of logit equations. In these models, discrete endogenous variables are not subject to consistent treatment. In partic-

SOC7912648 ("Social and Demographic Sources of Change in the Youth Labor Force"). Computations were performed on the VAX 11/780 at the University of Wisconsin Center for Demography and Ecology, supported by grant HD05876-11 from the National Institute of Child Health and Human Development. We are grateful to Ann Kremers and Warren Kubitschek for research assistance, to William Bielby for helpful comments on an earlier draft of this paper, to William Landes for providing us his data on state characteristics, and to David Wise for giving us his program for bivariate probit analysis. Requests for reprints should be sent to Christopher Winship, Department of Sociology, Northwestern University, Evanston, Illinois 60201.

[2] Davis (1975) proposes a path analysis for discrete data based on the linear probability model. Unlike logit models and the other approaches discussed here, however, this approach assumes that the effects of independent variables are constant over the entire (0,1) range of the dependent variable The shortcomings of this model are well-known (Hanushek and Jackson 1977, pp. 183–86).

ular, when a discrete intervening variable is the dependent variable, it appears as the logit of a probability; when it is an independent variable, it appears as a dummy variable. For example, consider a simple model in which social background factors affect whether or not an individual has a college degree, and background and degree status affect whether or not the individual experiences unemployment. As a dependent variable, degree status is measured as the logit transformed probability of having a degree; as an independent variable, it is a simple dummy variable. Although suitable in many applications, this treatment of discrete variables may not be suitable in others. In some applications it is, at best, an arbitrary formulation dictated by the use of log-linear models and not by its reasonableness in substantive applications.

Discrete variables arise in a variety of contexts. Some measure inherently discrete characteristics, for example, sex or race. Some arise out of imperfections in measurement instruments. For example, family income, a continuous variable, may be measured in survey data as the discrete variable, "less than or greater than $20,000 per year." Some measure inherently discrete observations that are of interest less in their own right than as indicators of latent continua. For example, whether or not a white respondent would be willing to live in a predominantly black neighborhood is a binary response that derives its substantive interest from its indication of underlying racial tolerance. Finally, some discrete variables may play dual roles; that is, they can act both as direct measurements of discrete phenomena and as indicators of unmeasured phenomena that are best conceived of as continuous. For example, whether or not a state or nation has a particular type of law is clearly discrete, but this fact may index the accumulated sentiments of populations or of decision-making bodies. In some analyses both the law itself and the sentiments it reflects may be of interest.

These alternative contexts in which discrete variables arise require analysts to select statistical models that are in accord with the kinds of effects implied by substantive reasoning. In many contexts, recently developed methods for the causal analysis of discrete data embody the substantive ideas of interest satisfactorily. In others, however, alternative models may be more appropriate.

An issue related to the proper formulation of causal models for discrete data is the capacity of such models to allow the application of path analysis procedures to discrete data. Structural equation models for continuous data allow the analyst to quantify the direct and indirect effects of predetermined variables on endogenous variables. In discrete systems, analyzed by existing methods, however, the inconsistent treatment of intervening variables implies that the usual "theorem of path analysis" (Duncan 1966) cannot be applied directly. This has led some commentators to conclude that "there is no calculus of paths" for these models (Fienberg 1980, p. 120).

This article presents several alternative approaches to the formulation of structural equation models for discrete data that take into account the variety of roles played by discrete variables in quantitative analysis. It shows that the current approach to causal analysis of discrete data is one of several that may be applied, depending on the underlying causal process that the analyst hypothesizes. In addition, the article shows that the difficulties encountered in applying path analytic principles have attractive solutions regardless of which approach to the analysis of discrete systems is adopted. It also shows that structural equation models for discrete data, though sometimes computationally complex, are every bit as flexible analytically as corresponding models for continuous variables. Indeed, with little difficulty, discrete and continuous response variables can be analyzed together in the models discussed here.

The insights from several literatures are synthesized in this study. There is an examination of alternative formulations of discrete variables when they are dependent variables; that is, either as the outcomes of inherently discrete stochastic processes (Yule 1900) or as realizations of latent continuous variables (Pearson 1900).[3] There is also consideration, based mostly on recent work by Heckman (1978), of the alternative formulations of the role of discrete variables when they are independent variables; that is, either as observed dummy variables or as indicators of unobserved continuous variables. In incorporating these alternative conceptions into structural equation models, the article applies the recent literature on simultaneous probit models (Zellner and Lee 1965; Ashford and Sowden 1970; Heckman 1978; Amemiya 1974, 1976, 1978; Muthén 1979) and on discrete choice (Hausman and Wise 1978; Manski 1981; McFadden 1976, 1980). Finally, it is shown that the problem of calculating direct and indirect effects in structural equation models can be solved by straightforward application of methods proposed by Stolzenberg (1979) for the analysis of nonlinear models. The contribution of this article, therefore, is to apply the contributions of diverse literatures to a central methodological problem in sociology.

This study is limited in two respects. First, only recursive structural equation models are considered. The ideas presented here can be extended to systems embodying simultaneity, but the latter are beyond the scope of this article. Discussions of nonrecursive systems are available elsewhere (Heckman 1976, 1978; Brier 1978; Zellner and Lee 1965). In any case, it is from recursive models that structural equation models have thus far borne their greatest fruit in sociology (e.g., Duncan, Featherman, and

[3] The possibility that discrete variables are indicators of latent categorical variables ("latent classes") is not considered here (Lazarsfeld and Henry 1968; Goodman 1974, Clogg 1980).

Duncan 1972; Sewell and Hauser 1975). Second, the treatment of discrete variables presented here is limited to variables with only two categories. Models with ordered response variables can, however, be developed from the ideas for dichotomous responses presented here (e.g., Amemiya 1981; McKelvey and Zavoina 1975).

The balance of the article is divided into four major sections. Section I discusses the problem of conceptualizing discrete variables when they are embedded in structural equation models. First it discusses discrete variables both as dependent variables and as independent variables, then it presents four alternative recursive models that embody distinct relationships between a discrete variable and other variables in the model. Section II discusses the identification of recursive models with discrete data. Section III shows how the methods of path analysis can be adapted to models with discrete variables. Section IV presents two numerical examples that illustrate the methods exposited in Sections I, II, and III. Methods of estimating the models discussed in the article are discussed in the Appendix.

I. THE CONCEPTUALIZATION OF DISCRETE VARIABLES

This section presents alternative formulations for discrete variables, first as dependent variables and then as independent variables. Then it applies these formulations to four simple recursive models containing discrete variables.

1. Discrete Variables as Dependent Variables

Historically, there have been two approaches to modeling discrete dependent variables: they have been modeled as observed indicators of unobserved continuous variables and as inherently discrete outcomes of binomial trials.

Discrete variables as realizations of underlying continuous variables: threshold model.—Discrete variables may be thought of as indicators of continuous variables that are either difficult or impossible to measure directly. Let d_y be a dichotomous variable with the value 1 or 0 and let Y^* be an unobserved continuous variable. Here and throughout the article, unobserved variables such as Y^* will be assumed to have means of zero and unit variance. Alternative scaling assumptions do not materially affect the models and methods presented. Under this model, Y^* and d_y are related as follows:

$$\begin{aligned} d_y &= 1 \text{ if } Y^* \geqslant L \,, \\ d_y &= 0 \text{ if } Y^* < L \,; \end{aligned} \tag{1}$$

where L is a parameter denoting the "threshold" across which d_y changes from 0 to 1. Thus Y^* and d_y are related by a nonlinear transformation: all values of Y^* above or equal to L have been transformed to 1; all values below L have been transformed to 0. There are numerous instances where this formulation is applicable. An attitude item on whether or not the respondent is willing to live in a predominantly black neighborhood is a binary outcome that may index whether the respondent is above or below a threshold on an unmeasurable continuum of racial tolerance/prejudice. Data obtained on whether or not an individual has attended college may index the approximately continuous variable, length of time in school, which was not obtained in a survey or published tabulation. If Y^* is affected by an independent variable, say X, then a model may be written as

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y \, , \tag{2}$$

where β_0 and β_1 are parameters to be estimated, ϵ_y is a stochastic disturbance assumed to be uncorrelated with X, and Y^* is related to the observed binary variable as in (1) above.[4] The disturbance ϵ_y is assumed to follow a probability distribution that remains to be specified. If ϵ_y has a normal distribution, (1) and (2) define a probit model (Hanushek and Jackson 1977, pp. 204–5). If ϵ_y has an extreme value distribution (Johnson and Kotz 1970, pp. 272–89), (1) and (2) define a logit model (McFadden 1974). The logit and probit models are discussed in more detail in the Appendix.

Figures 1*A* and *B* show graphically the essential features of this model for discrete response variables. Figure 1*A* shows that the dichotomous variable is an exact step function of Y^*. Figure 1*B* provides a path diagram of the model. The wavy line connecting Y^* and d_y denotes that the relationship is nonlinear and deterministic; the solid line connecting X and Y^* denotes a linear, stochastic relationship.

Discrete variables as outcomes of binomial trials.—A second model for binary data is that observed dichotomous variables arise from binomial trials; that is, that the phenomena of interest are inherently discrete and observed binary outcomes are stochastic in nature. This can be written

[4] Here and elsewhere models are presented with a single independent variable X. Adding additional independent variables to the models raises no new issues beyond those discussed for single independent variables. An alternative equivalent specification of the threshold model is to regard the threshold L as random, rather than fixed, and the latent continuous variable Y^* as a deterministic rather than stochastic function of X. That is, $L = \bar{L} + \epsilon_L$, and $Y^* = \beta_0 + \beta_1 X$; where L is now a random threshold, $\bar{L}$ is the mean of L, and ϵ_L is a stochastic disturbance. If $\epsilon_L = \epsilon_y$, this model is equivalent to the one discussed above. Although the model presented here postulates a stochastic relationship between d_y and Y^*, it differs from the stochastic model presented below in which Y^* is not only stochastically related to d_y but also a function of X and a stochastic disturbance (see n. 7 below). To simplify the exposition, the random threshold formulation is not adopted in the text.

$$d_y = 1 \text{ with probability } p\,, \qquad d_y = 0 \text{ with probability } 1 - p\,; \tag{3}$$

where p is the unobserved probability which governs the distribution of 0's and 1's according to some mechanism (e.g., the flip of a coin). This formulation underlies most of the models in discrete data analysis as applied to binary dependent variables (Cox 1970; Fienberg 1980) and is suited to such examples as whether an individual is employed, whether he or she dies, or whether he or she was a robbery victim. In these examples, the event—unemployment, death, or robbery—is defined relative to a fixed period, for example, a year. If an independent variable

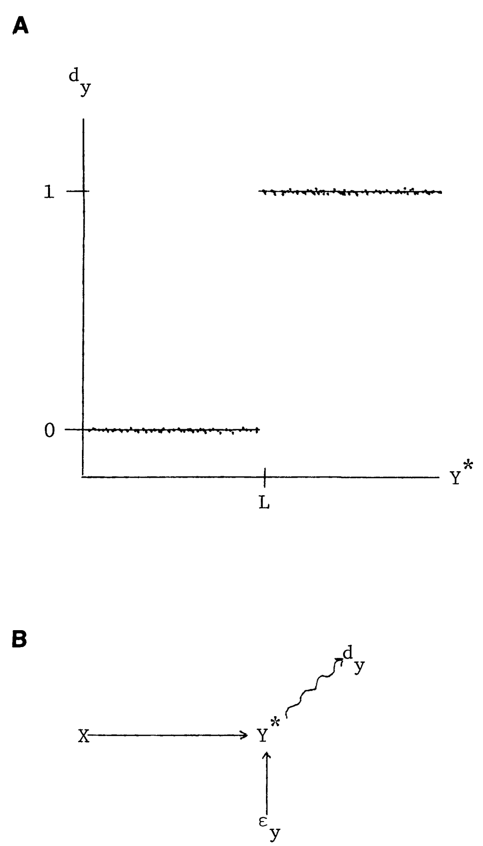

FIG. 1.—Threshold model. *A*, Relationship between observed d_y and unobserved Y^*. *B*, Path diagram.

X affects the binary outcome, a model can be specified to allow X to affect the probability that the outcome is a 1 (or a 0). That is,

$$p = F(\beta_0 + \beta_1 X) , \tag{4}$$

where F is the cumulative density (distribution) function for some symmetric probability distribution that must be specified and β_0 and β_1 are parameters.[5] If F is the logistic function, (4) is a logit model.[6] If F is the cumulative normal density function, (4) is a probit model. An alternative, but equivalent, way of expressing the effect of X on the discrete outcome is to define an unobserved variable Y^* which is a transformation of the unobserved probability p such that it is a linear function of X and an error term ϵ_y. That is,

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y , \tag{5}$$

with

$$\begin{aligned} d_y &= 1 \text{ with probability } G(Y^*) , \\ d_y &= 0 \text{ with probability } 1 - G(Y^*) , \end{aligned} \tag{6}$$

where G is the cumulative density function of a distribution that remains to be specified. Then the standard binomial trials model is obtained by assuming that ϵ_y is zero for all observations, that is, that all observations with the same X have equivalent probabilities that $d_y = 1$ (Amemiya and Nold 1975; Hanushek and Jackson 1977, pp. 99–200, 203). As is demonstrated below, without this assumption the model is not identified.

Figures 2*A* and *B* represent the essential features of the binomial trials model as formulated by equations (5) and (6). Figure 2*A* shows the relationship between the transformation of the unobserved probability and the observed dummy variable. In contrast to the threshold model where the continuous and discrete variables are related by a nonlinear but deterministic function, in the binomial trials model, Y^* and d_y are nonlinearly and stochastically related. Figure 2*B* presents a path diagram of the effect of X on d_y. The dotted line denotes the nonlinear stochastic relationship plotted in figure 2*A*.

Relationship between the two models.—The binomial trials and thresh-

[5] For probability distributions of general shape,

$$p = \int_{-(\beta_0+\beta_1 X)}^{\infty} f(u)du = 1 - F[-(\beta_0 + \beta_1 X)] ,$$

where f is a probability density function and F is its corresponding distribution function. For symmetric F, this expression simplifies to (4).

[6] The logistic function is the cumulative density function of the extreme value distribution (Johnson and Kotz 1970, pp. 272–89).

old models are related as follows. Let Y' be a continuous unobserved variable with mean 0 and variance 1 such that

$$d_y = 1 \text{ if } Y' \geqslant M , \quad d_y = 0 \text{ if } Y' < M . \tag{7}$$

Then the binomial trials model may be written

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y , \tag{8}$$

$$Y' = \lambda Y^* + \epsilon_{y'} , \tag{9}$$

where λ is the zero-order coefficient for the regression of Y' on Y^* and $\epsilon_{y'}$ is a stochastic disturbance. In this formulation, there is a variable Y' which has a threshold M dividing observations with $d_y = 1$ from those

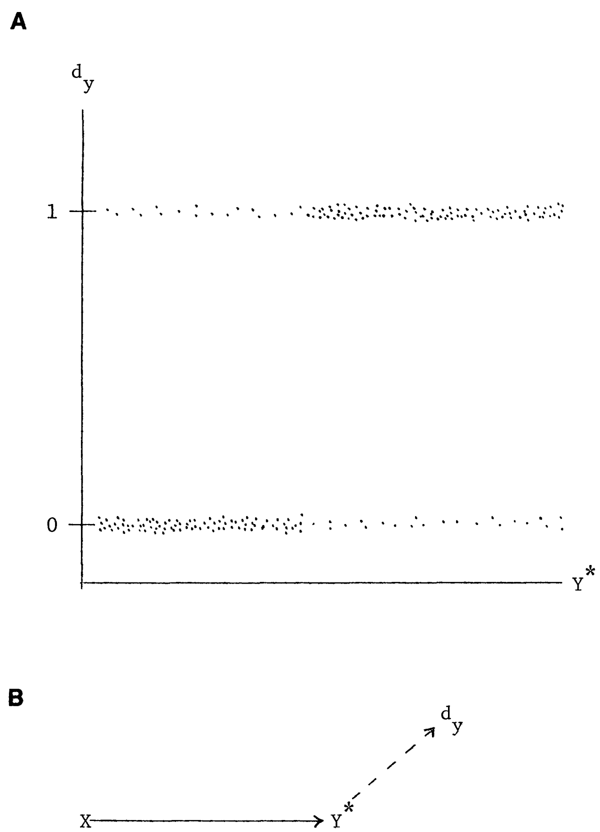

FIG. 2.—Binomial trials model. *A*, Relationship between d_y and Y^*. *B*, Path diagram.

with $d_y = 0$. The variable Y', however, is not itself a direct linear function of the independent variable X but instead is affected linearly by Y^* which, as before, is a transformation of the unobserved $p(d_y = 1)$ and is a stochastic linear function of X. Equations (8) and (9) are a structural equation and a measurement equation, respectively. Thus d_y can be interpreted as a discrete variable with measurement error, which imperfectly classifies observations into whether Y^* is less or greater than M. The threshold model is a special case of equations (7), (8), and (9) where $\lambda = 1$ and $\epsilon_{y'} = 0$. Thus the threshold model is a restricted case of the binomial trials model that assumes no measurement error in d_y. In practice, it may make no difference whether or not λ can be separated from the structural relationship between X and Y^*.[7]

In the past, the relative merits of the threshold and binomial models of discrete variables have occasioned considerable debate, notably between Yule, who advocated the binomial trials approach, and Pearson, who advocated the threshold approach (Fienberg 1980). In recent decades, this debate has translated into the parallel lines of intellectual development of log-linear models for discrete data, a method which has assumed a binomial trials model, and biometric and econometric approaches (Ashford and Sowden 1970; Zellner and Lee 1965) which have assumed a threshold model. As will be shown below, each approach has considerable applicability to the analysis of discrete data in the social sciences depending on the substantive context and the purposes of the analyst (Goodman 1981).

2. Discrete Variables as Independent Variables

The preceding discussion showed that discrete variables may arise either as indicators of unobserved continuous variables that may be of greater substantive interest or as results of unobserved (continuous) probabilities (or their transformations) that lack substantive content. In both instances, however, they arise as continuous variables that are linearly related to independent variables. In contrast, when discrete variables are treated as independent variables, they may occupy two distinct roles. That is, they may measure discrete phenomena of direct substantive interest or they may be indicators of unobserved continuous variables.[8]

[7] An alternative way to relate these two models is to assume that, in the threshold model, the thresholds are random, rather than fixed; i.e., $L = \bar{L} + \epsilon_L$ (see n. 4 above). This model can be shown to be equivalent to the binomial trials model as follows: substitute (9) into (7) to obtain $d_y = 1$ if $Y^* \geq (M - \epsilon_{y'})/\lambda$, $d_y = 0$ if $Y^* < (M - \epsilon_{y'})/\lambda$. Then the two models are equivalent if $\bar{L} = M/\lambda$ and $\epsilon_L = \epsilon_{y'}/\lambda$. Note that this model generalizes the random threshold model discussed in n. 4 above by relaxing the assumption made in the latter model that Y^* is an exact linear function of X.

[8] This discussion draws much from Heckman (1978).

Discrete variables as indicators of discrete phenomena.—A discrete variable may have a direct effect on a dependent variable. In this case it enters a linear equation as a dummy variable denoting that the group or characteristic of interest has a different intercept from other observations in the analysis, that is, there is a "structural shift" associated with the group or characteristic (Heckman 1976, 1978). If Z is an endogenous variable (which may be either an observed continuous variable or an unobserved continuous variable with discrete indicators), this type of effect of a discrete variable is denoted by α_2 in the equation $Z = \alpha_0 + \alpha_1 X + \alpha_2 d_y + \epsilon_z$, where α_0, α_1, and α_2 are parameters, X is an independent variable, ϵ_z is a stochastic disturbance, and, as before, d_y is a dummy variable taking the value 1 or 0.

Discrete variables as indicators of unobserved continuous variables.—Alternatively, an observed discrete variable itself may exert no effect on a dependent variable but instead may be an indicator of an unobserved continuous variable that affects the dependent variable. Such a model can be written $Z = \alpha_0 + \alpha_1 X + \alpha_3 Y^* + \epsilon_z$, where $d_y = 1$ with probability $F(Y^*)$, $d_y = 0$ with probability $1 - F(Y^*)$, and all notation is as defined above.

These two models reflect distinct conceptions of how a discrete variable exerts its effect. In the first case, the effect of the discrete variable is direct; in the second, the observed variable is merely an indicator for the continuous variable of interest. Consider several examples of this distinction.

1. Political party membership, indicated by a discrete variable (Republican versus Democrat), may affect an individual's attitude toward proposed legislation. If the legislation is proposed by an individual's own party and party leadership is followed, the discrete variable has a direct positive effect on the individual's attitude. Alternatively, an individual may be a party member but support legislation because of her or his ideological beliefs. In this case, there is no structural relationship between membership and attitude toward legislation. However, a continuous, unobserved variable, ideology, of which party membership *may* be an indicator, might in itself affect attitude toward legislation.

2. Marital disruption may affect the chances that the children of separated spouses will drop out of high school. Whether or not parents have separated (a discrete variable) may affect dropout chances. Alternatively, the separation of parents may index family conflict or disharmony (an unobserved variable), which may exert more of an effect on the probability of dropping out. These alternative conceptions imply distinct models for the discrete variable, marital disruption. Indeed, they harbor distinct implications for the desirability of parents in conflict-ridden marriages staying together until their children have completed school.

3. Whether or not an infant has a low birth weight has a pronounced effect on its chances of survival (Gortmaker 1979). Measuring "low birth weight" as a dichotomy, low weight or not low weight, implies that there is a critical weight below which an infant's chances of death increase astronomically. It may, however, be more reasonable to assume that the relationship between birth weight and mortality is continuous (albeit nonlinear) and that the dichotomous measure should be regarded as an indicator of a continuous measure of weight.

These examples illustrate that the way in which discrete variables should be treated as independent variables in structural equation models is closely linked to substantive arguments about the way their effects are hypothesized to occur.

3. Some Recursive Models

The foregoing ideas can be combined to yield recursive models which reflect alternative assumptions about the role of discrete variables. Four models are considered here. In each a single, predetermined variable X, which may be either discrete or continuous, affects an intervening variable Y, which is observed as a discrete binary variable d_y, and both X and Y affect a third variable Z, which may be discrete or continuous. These models can be extended to include additional predetermined and intervening variables. The models are presented here and their identifiability is discussed in the next section.

Model I: intervening variable as unobserved continuous variable deterministically related to observed discrete variable.—Consider first the model represented by the path diagram in figure 3. The model may be written

$$Z = \alpha_0 + \alpha_1 X + \alpha_2 Y^* + \epsilon_z ,$$

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y ;$$

where

$$\text{cov}(\epsilon_z,\epsilon_y) = 0 ,$$

$$d_y = 1 \text{ if } Y^* \geqslant L ,$$

$$d_y = 0 \text{ if } Y^* < L ,$$

and all notation is as defined above. In this model the observed discrete variable d_y does not enter the two structural equations of the model. Instead it is an indicator of an unobserved variable Y^*, to which it is deterministically related as in figure 1*A*. As will be shown below, it is possible to identify and estimate the structural parameters α_0, α_1, α_2, β_0, and β_1.

Model II: intervening variable as unobserved continuous variable stochastically related to observed discrete variable.—Figure 4*A* represents model II in which the discrete variable d_y is an indicator for an unobserved continuous variable as in model I. Here, however, d_y is the outcome of binomial trials, and thus Y^* and d_y are related stochastically as in figure 2*A*. The model may be written

$$Z = \alpha_0 + \alpha_1 X + \alpha_2 Y^* + \epsilon_z ,$$
$$Y^* = \beta_0 + \beta_1 X + \epsilon_y ;$$

where

$$\text{cov}(\epsilon_z, \epsilon_y) = 0 ,$$
$$p(d_y = 1) = F(Y^*) ,$$
$$p(d_y = 0) = 1 - F(Y^*) ,$$

and, as above, F is a cumulative probability function such as the logistic or cumulative normal.

An equivalent, alternative formulation of model II can be obtained by defining a continuous unobserved variable Y' which is linearly (though not exactly) related to Y^* and is nonlinearly but deterministically related

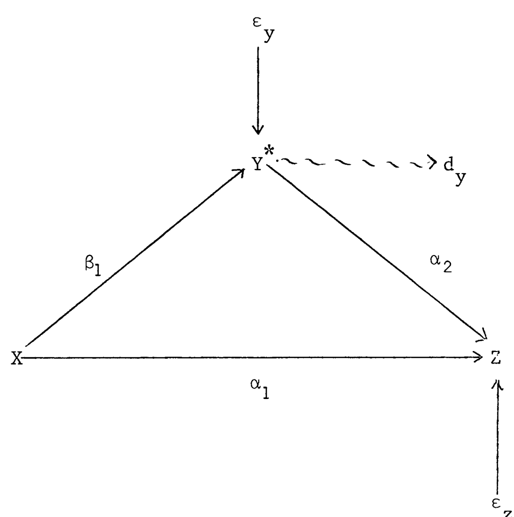

FIG. 3.—Path diagram of model I

to d_y. Thus, as in the preceding section, Y' is a variable which is classified perfectly by d_y, just as Y^* is classified perfectly by d_y in model I. Model II may then be represented as in figure 4*B*, which shows that Y^* and Y' are related by the measurement equation $Y' = \lambda Y^* + \epsilon_{y'}$, where $d_y = 1$ if $Y' \geqslant M$, $d_y = 0$ if $Y' < M$. This formulation shows that model II contains one more parameter (λ) than model I, indicating that the relationship between the unobserved continuous variable and its discrete indicator d_y is not exact and needs to be estimated from the data. As will be shown below, the parameters of model II are not identifiable, in gen-

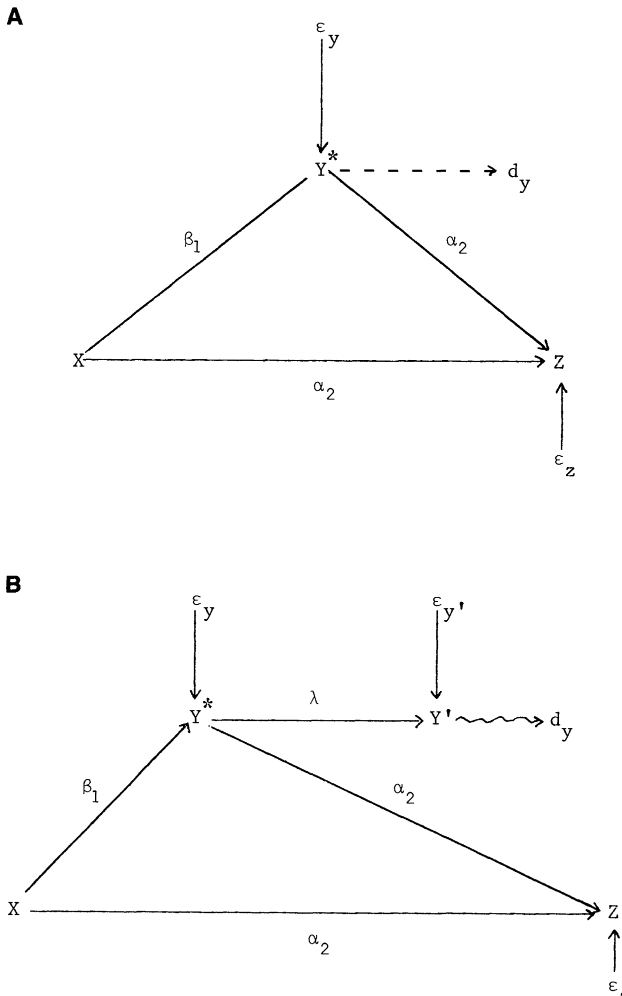

FIG. 4.—Model II. *A*, Path diagram of model II. *B*, Path diagram of model II: threshold representation.

eral, without additional information. This information, however, is often available, and simple modifications of model II are identifiable.

Model III: intervening variable as observed discrete variable deterministically related to unobserved continuous variable.—In contrast to the two models discussed above, in the third and fourth models the observed discrete intervening variable d_y affects Z directly. In model III, d_y and an unobserved continuous variable Y^*, which is a linear function of X, are related deterministically; thus this model parallels model I. Model III is diagramed in figure 5. The model is

$$Z = \alpha_0 + \alpha_1 X + \alpha_3 d_y + \epsilon_z ,$$

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y ;$$

where

$$\text{cov}(\epsilon_z, \epsilon_y) = 0 ,$$

$$d_y = 1 \text{ if } Y^* \geq L ,$$

$$d_y = 0 \text{ if } Y^* < L .$$

Figure 5 indicates that model III differs from typical three-variable recursive models because it contains an unobserved variable Y^* which is related nonlinearly to an observed variable d_y but (in contrast to models I and II) both variables enter the structural equations. The structural

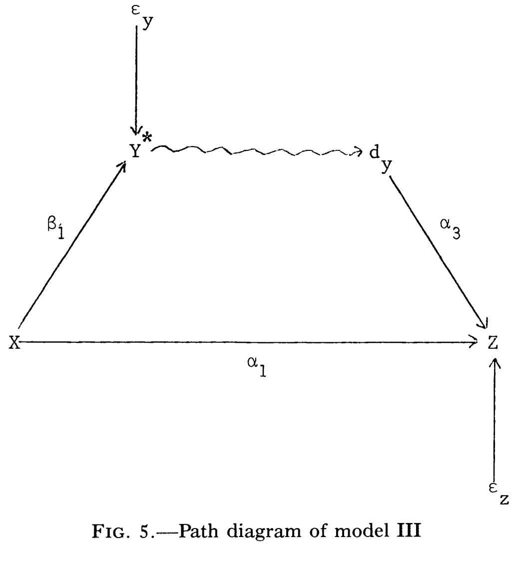

FIG. 5.—Path diagram of model III

parameters α_0, α_1, α_3, β_0, and β_1 are nonetheless identifiable, as will be shown in the next section.

Model IV: intervening variable as observed discrete variable stochastically related to unobserved continuous variable.—In model IV, d_y and an unobserved continuous variable Y^*, which is a linear function of X, are related stochastically. Thus this model parallels model II. That is, d_y is the outcome of binomial trials. Model IV is diagramed in figure 6*A*

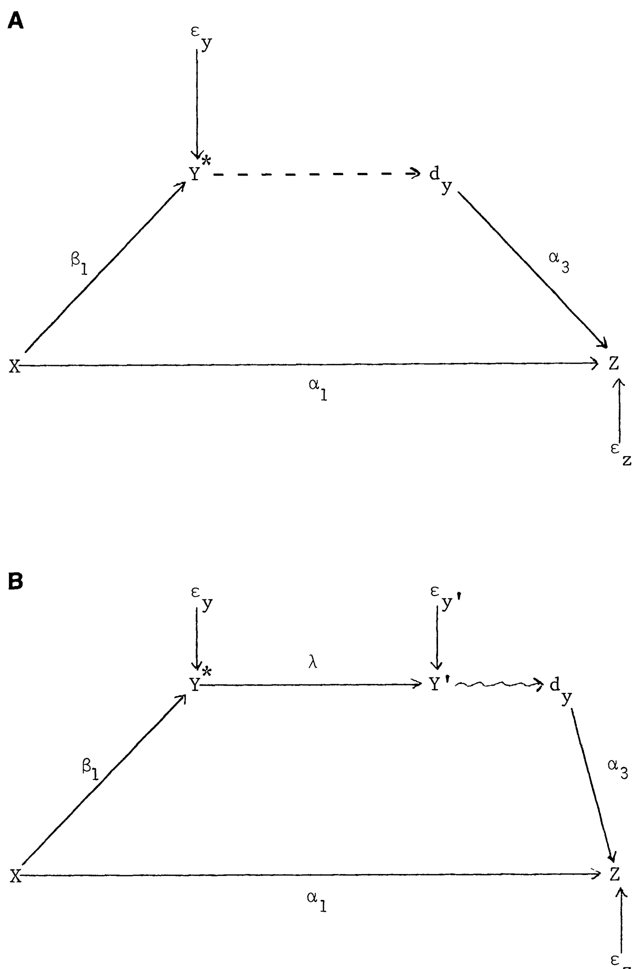

FIG. 6.—Model IV. *A*, Path diagram of model IV. *B*, Path diagram of model IV: threshold representation.

and may be written

$$Z = \alpha_0 + \alpha_1 X + \alpha_3 d_y + \epsilon_z \ ,$$

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y \ ;$$

where

$$\text{cov}\ (\epsilon_z, \epsilon_y) = 0 \ ,$$

$$p(d_y = 1) = F(Y^*) \ ,$$

$$p(d_y = 0) = 1 - F(Y^*) \ .$$

Like model II, model IV can be reformulated to show explicitly the stochastic relationship between Y^* and d_y by introducing the continuous unobserved variable Y', which is related deterministically to d_y. Figure 6*B* shows this alternative formulation, and the equations of the model can be augmented to include the measurement equation $Y' = \lambda Y^* + \epsilon_{y'}$, where $d_y = 1$ if $Y' \geqslant M$, $d_y = 0$ if $Y' < M$. Model IV contains one more parameter (λ) than model III, reflecting the stochastic relationship between Y^* and d_y. As will be shown in the next section, model IV is not in general identifiable without additional information, although it can be identified in a wide range of circumstances.

The formal properties of models I–IV discussed thus far are summarized in the first two columns of table 1. Substantive differences among the models can be appreciated from two examples.

Consider a simple recursive model in which an individual's parents' socioeconomic status influences his grades of formal schooling achieved and both socioeconomic status and schooling affect earnings. Suppose, however, that the analyst does not measure schooling as a continuous variable but measures only whether or not individuals graduated from high school. This binary measure of schooling may capture the variation in schooling that affects earnings. Either the credential of a high school degree or the human capital derived from completion of all high school requirements may be the primary determinant of earnings (e.g., Jencks et al. 1979). Alternatively, however, the effect of grades of schooling achieved may be linear, with no bonus derived from high school completion. Given the data, these alternative hypotheses could be formulated using models I and III above. In figure 3, let Z be earnings, X be socioeconomic status, Y^* be an unmeasured variable interpreted as a continuous measure of schooling, and d_y be the observed dichotomy, graduate versus nongraduate. Under model I, schooling affects earnings as an unobserved continuous variable. Under the alternative hypothesis (model III, fig. 5), the variables denote the same concepts, but schooling affects earnings as an observed discrete variable.

A second example is the task of determining whether or not prior robbery victimization affects the likelihood of gun ownership when other

TABLE 1

PROPERTIES OF ALTERNATIVE RECURSIVE MODELS WITH DISCRETE INTERVENING VARIABLE (Y^*)

Model	Relationship between Discrete Outcome (d_y) and Unmeasured Continuous Variable (Y^*)	Causal Form of Intervening Variable	Identifiable from Simple Single-Equation Estimates	Identifiable from Reduced-Form Error Correlation	Identifiable with Instrumental Variables	Identifiable with Multiple Indicators of Y^*
I	Deterministic	Unobserved continuous variable (Y^*)	No	Yes	Yes	Yes
II	Stochastic	Unobserved continuous variable (Y^*)	No	No	Yes	Yes
III	Deterministic	Observed dummy variable (d_y)	Yes	Yes	Yes	Yes
IV	Stochastic	Observed dummy variable (d_y)	No	No	No	Yes

factors correlated with victimization, such as the crime rate in an individual's neighborhood, may also be important causes. If whether or not an individual owns a gun, whether or not he has recently been a victim, and other personal background factors are known, but neighborhood factors are unmeasured, then model II can be used to represent the alternative mechanisms through which victimization might affect ownership; to wit, either directly or as an indicator of the neighborhood climate that the individual experiences. In figure 4*A* and *B*, let X denote measured personal background, Z denote gun ownership, Y^* denote the neighborhood climate experienced by the individual, and d_y denote whether or not the individual has recently been a robbery victim. Under this model, gun ownership is affected by a latent continuous variable that is indicated by an observed discrete variable. In contrast, model IV (fig. 6*A* and *B*) represents a direct effect of observed victimization on ownership. In this example, gun ownership is a discrete variable. Strictly speaking, figures 4 and 6 should be modified to indicate that Z is an unobserved continuous variable related linearly to the predetermined variables and related nonlinearly to a discrete variable, say d_z. This modification, however, would leave unchanged the logical issues involved in considering the effects of Y^* and d_y. Models I and III might also be used to explore these arguments. Insofar as models II and IV are identifiable, however, they are most likely preferable. The latter models allow the relationship between neighborhood climate and personal experience to be stochastic. That is, some residents of dangerous neighborhoods have not been victims, and conversely, some residents of safe neighborhoods have been victims. Models I and III, in contrast, assume that whether or not an individual has been a victim exactly differentiates neighborhoods by their safety.

These examples raise the issue of testing rigorously among the alternative models. In practice, this requires that a more general model, one in which both the continuous and discrete effects of Y are estimated, be compared to the two more restrictive models (either I and III or II and IV). This procedure is illustrated in Section IV.

Of the four models considered in this section, model IV is closest to the recursive systems of logit equations for the causal analysis of discrete data discussed by Goodman (1972, 1973*a*, 1973*b*, 1979) and others. In these systems, when the discrete intervening variable is a dependent variable, it is formulated as the logit of the unobserved probability of an outcome on the discrete variable.[9] When the discrete variable is an in-

[9] The models discussed by Goodman (e.g., 1979) are equivalent to model IV if $\epsilon_y = 0$ and Y^* is viewed as a transformed unobservable probability p that has no substantive content beyond that contained in p itself. Model III can also be estimated as a log-linear or logit model, but here Y^* is viewed as an unobserved continuous variable with discrete indicator d_y. This is a different formulation from that of Goodman.

dependent variable, it is a dummy variable. Thus these methods for analysis of recursive systems are best applied to the substantive contexts for which model IV is appropriate.[10]

II. MODEL IDENTIFICATION

This section discusses the identification of recursive models for discrete data typified by models I–IV. First it examines the problem of identifying parameters in these models, focusing initially on single-equation models for discrete response variables and then on models I–IV in turn. The discussion of model identification assumes that equations with continuous response variables can be estimated by ordinary least squares (OLS), that equations with binary response variables can be estimated by logit or probit analysis, and that models with two or more equations can be estimated by extensions of these procedures. These methods of estimation are discussed in the Appendix.

Identification of parameters in structural equation models with discrete variables comprises two parts, the identification of parameters in a single equation with a continuous unobserved dependent variable indicated by an observed dichotomous variable, and the identification of parameters in multiequation models.

1. Single-Equation Models with Binary Dependent Variables

The preceding section discussed models in which discrete variables are indicators of unobserved continuous dependent variables, a formulation that was termed the "threshold model." As before, let d_y be a discrete response variable; X and W be observed independent variables; β_0, β_1, and β_2 be parameters; Y^* be an unobserved continuous variable; and ϵ_y be a stochastic disturbance with variance $\sigma_{\epsilon_y}^2$ assumed to be uncorrelated with X or W. Following the earlier discussion of the threshold model, $Y^* = \beta_0 + \beta_1 X + \beta_2 W + \epsilon_y$, where $d_y = 1$ if $Y^* \geqslant L$, $d_y = 0$ if $Y^* < L$, and L is the threshold parameter.

Without further specification, this model is underidentified because the scale of the unobserved variable Y^*, and thus of ϵ_y, is unknown. Under

[10] In practice, issues of identifiability and the treatment of the stochastic disturbances in these models are not addressed explicitly. Typically, it is assumed in the notation of model IV that $\lambda = 1$ and $\epsilon_y = 0$. Given the binomial trials interpretation that underpins such models, however, the parameters estimated under these systems of logit equations for the effect of X on Y are best interpreted as $\lambda\beta$, i.e., the product of the structural and measurement parameters. With sufficient information, however, the measurement and structural parts of model IV can be identified.

either a logit or probit specification, the estimated model is $y = b_0 + b_1X + b_2W + e_y$, where $y = Y^*/\sigma_{\epsilon y}$, $b_0 = \beta_0/\sigma_{\epsilon y}$, $b_2 = \beta_2/\sigma_{\epsilon y}$, and $e_y = \epsilon_y/\sigma_{\epsilon y}$. To identify the structural parameters β_0, β_1, β_2, and $\sigma_{\epsilon y}$, an assumption about the variance of Y^* or ϵ_y is required. For example, if $\sigma_{\epsilon y}^2 = 1$, then $\beta_0 = b_0$, $\beta_1 = b_1$, and $\beta_2 = b_2$. An alternative scaling assumption, employed throughout this article, is that var $(Y^*) = 1$. Under this assumption,

$$\begin{aligned} 1 &= \beta_1^2 \text{ var } (X) + \beta_2^2 \text{ var } (W) + 2\beta_1\beta_2 \text{ cov } (X,W) + \sigma_{\epsilon_v}^2 \\ &= \sigma_{\epsilon_y}^2 b_1^2 \text{ var } (X) + \sigma_{\epsilon_y}^2 b_2^2 \text{ var } (W) + 2\sigma_{\epsilon_y}^2 b_1 b_2 \text{ cov } (X,W) + \sigma_{\epsilon_y}^2 , \end{aligned}$$

and thus the parameters of interest can be obtained from the estimated parameters with the equations

$$\sigma_{\epsilon_y} = \sqrt{1/[b_1^2 \text{ var } (X) + b_2^2 \text{ var } (W) + 2b_1b_2 \text{ cov } (X,W) + 1]} ;$$

$$\beta_0 = b_0\sigma_{\epsilon_y} ; \quad \beta_1 = b_1\sigma_{\epsilon_y} ; \quad \beta_2 = b_2\sigma_{\epsilon_y} .$$

In contrast to models with observed continuous dependent variables, therefore, a scaling assumption is required to identify model parameters.[11] With the scaling assumption used here, the slope parameters β_1 and β_2 measure the effects of a unit change in the independent variables X and W measured in standard deviations of the latent variable Y^*. Similar arguments are necessary to identify parameters in multiequation models discussed below.[12]

The binomial trials model presents identification problems in addition to that of scale identification. Consider again the threshold representation of the binomial trials model, that is, as in equations (8) and (9) above: $Y^* = \beta_0 + \beta_1X + \epsilon_y$, $Y' = \lambda Y^* + \epsilon_{y'}$, where $d_y = 1$ if $Y' \geq M$, $d_y = 0$ if $Y' < M$. Because this model contains an additional source of variation, namely, $\epsilon_{y'}$, and an additional parameter λ, assumptions about the variances Y^* and Y' are insufficient to identify the model. Assume that both Y^* and Y' have variances of unity. Substitution of (8) into (9) yields $Y' = \lambda\beta_0 + \lambda\beta_1X + \lambda\epsilon_y + \epsilon_{y'}$. Because this equation can be estimated

[11] This discussion has used the notation of the threshold model (eqq. [1] and [2]). The same issue of scale identification arises in the binomial trials model (eqq. [3]–[9]), although the additional parameter λ in the latter model raises additional identification problems (see below).

[12] A further requirement for the identification of single-equation models is that the dichotomous dependent variable does not classify perfectly any of the independent variables or linear combinations of the independent variables; i.e., there can be no value of an independent variable above which all values of d_y are 1 and below which all values are 0 (Heckman 1978).

directly through logit or probit analysis, the quantities $\lambda\beta_0$ and $\lambda\beta_1$ are identified, but the separate structural and measurement parameters β_0, β_1, and λ are not.

If one assumes, however, that Y^* is an exact linear function of X, that is, that $\epsilon_y = 0$ for all observations, it is possible to identify both the structural and measurement parameters. Substitution of (8) into (9) assuming $\epsilon_y = 0$ yields $Y' = \lambda\beta_0 + \lambda\beta_1 X + \epsilon_{y'}$, which can be estimated by probit or logit analysis. As in the case of the threshold model, however, the estimated equation is $Y' = c_0 + c_1 X + e_{y'}$, where $y' = Y'/\sigma_{\epsilon y}$, $c_0 = \lambda\beta_0/\sigma_{\epsilon y'}$, $c_1 = \lambda\beta_1/\sigma_{\epsilon y'}$, $e_{y'} = \epsilon_{y'}/\sigma_{\epsilon y'}$, and $\sigma_{\epsilon y'}{}^2$, $=$ var $(\epsilon_{y'})$. If var $(Y') = 1$, then $\sigma_{\epsilon y'}{}^2 = 1/[1 + c_1{}^2 \text{ var } (X)]$ and

$$\lambda = 1 - \sqrt{1/[1 + c_1{}^2 \text{ var } (X)]} \; .$$

The structural parameters can be obtained as follows: $\beta_0 = c_0\sigma_{\epsilon y'}/\lambda$; $\beta_1 = c_1\sigma_{\epsilon y'}/\lambda$. Alternative strategies are discussed below. One means of identifying both the structural and measurement parts of the binomial trials model is to employ multiple indicators of Y^* (Muthén 1979). This approach is illustrated for model IV below. Another approach is to use additional information in the form of an instrumental variable. This is illustrated for model III.

2. Model I

Consider again the structural equations of model I:

$$Z = \alpha_0 + \alpha_1 X + \alpha_2 Y^* + \epsilon_z \; , \qquad (10)$$

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y \; ; \qquad (11)$$

where $d_y = 1$ if $Y^* \geq L$, $d_y = 0$ if $Y^* < L$, Y^* is a continuous unobserved variable, cov $(\epsilon_z, \epsilon_y) = 0$, var $(Y^*) = 1$, var $(\epsilon_z) = \sigma_{\epsilon z}{}^2$, and var $(\epsilon_y) = \sigma_{\epsilon y}{}^2$. Equation (11) can be estimated directly by probit or logit analysis, given the scale identification assumption on Y^*. Thus β_0, β_1, and $\sigma_{\epsilon y}{}^2$ are identifiable. Equation (10), however, creates additional problems because Y^* is unobserved and thus standard estimation procedures for either discrete or continuous measures of Z cannot be applied directly. Equation (10) is nonetheless identified, as can be seen by substituting (11) into (10) to yield the reduced-form equation

$$Z = \gamma_0 + \gamma_1 X + \eta_z \; , \qquad (12)$$

where

$$\gamma_0 = \alpha_0 + \alpha_2\beta_0 \; , \qquad (13)$$

$$\gamma_1 = \alpha_1 + \alpha_2\beta_1 , \tag{14}$$

$$\eta_z = \epsilon_z + \alpha_2\epsilon_y ,$$
$$\text{var}(\eta_z) = \sigma_{\eta_z}^2 = \sigma_{\epsilon_z}^2 + \alpha_2^2\sigma_{\epsilon_y}^2 , \tag{15}$$

and

$$\text{cov}(\eta_z,\epsilon_y) = \alpha_2\sigma_{\epsilon_y}^2 . \tag{16}$$

If Z is continuous, (12) can be estimated by OLS; if Z is discrete, (12) can be estimated by probit or logit analysis. As is shown in the Appendix, if an assumption is made about the multivariate distribution of ϵ_z and ϵ_y, and thus about η_z, it is possible to estimate cov (η_z,ϵ_y). Thus (13), (14), and (16) provide three equations in the three unknown parameters α_0, α_1, and α_2. The parameters of (10), therefore can be identified from the equations:

$$\alpha_0 = \gamma_0 - \beta_0 \left[\frac{\text{cov}(\eta_z,\epsilon_y)}{\sigma_{\epsilon_y}^2}\right] ,$$

$$\alpha_1 = \gamma_1 - \beta_1 \left[\frac{\text{cov}(\eta_z,\epsilon_y)}{\sigma_{\epsilon_y}^2}\right] ,$$

$$\alpha_2 = \frac{\text{cov}(\eta_z,\epsilon_y)}{\sigma_{\epsilon_y}^2} .$$

Thus, model I is identifiable from equations predicting Y^* and from the parameters of the reduced-form equation for Z.

This identification procedure requires that an estimate of the disturbance covariance, cov (η_z,ϵ_y), for (11) and (12) be obtained. Although this is always possible, it requires that one assume a particular multivariate distribution for ϵ_z and ϵ_y, and it is computationally burdensome.[13] In the absence of this estimate, model I cannot be identified. However, if there is additional information in the form of an instrumental variable for Y^*—that is, a variable, say W, that affects Y^* but does not affect Z directly—

[13] We are unaware of any research on the robustness under alternative distributional assumptions of the models considered here. As is well-known, maximum likelihood estimates are inconsistent if they rest on incorrect distributional assumptions, and it is usually preferable to adopt methods that are robust. Although it is applied to models quite different from those considered here, recent work by Heckman and Singer (1981) and by Goodman (1981) explores the implications of varying distributional assumptions and may lead to more robust methods.

then it is possible to identify the model without cov (η_z,ϵ_y). To see this, modify (11) as follows:

$$Y^* = \beta_0 + \beta_1 X + \beta_2 W + \epsilon_y . \tag{11*}$$

Then the reduced-form equation for Z (eq. [12]) becomes

$$Z = \gamma_0 + \gamma_1 X + \gamma_2 W + \eta_z , \tag{12*}$$

where

$$\gamma_2 = \alpha_2\beta_2 . \tag{17}$$

Equations (13), (14), and (17) contain the three parameters α_0, α_1, and α_2 and can be solved for them as follows: $\alpha_0 = \gamma_0 - (\beta_0\gamma_2/\beta_2)$; $\alpha_1 = \gamma_1 - (\beta_1\gamma_2/\beta_2)$; $\alpha_2 = \gamma_2/\beta_2$.

3. Model II

Model II, in which the observed discrete variable d_y is related stochastically to an unmeasured continuous variable, contains one more parameter than model I. As a result, without additional information it is not identified. To see this, consider the model, which is diagramed in figure 4*B:*

$$Z = \alpha_0 + \alpha_1 X + \alpha_2 Y^* + \epsilon_z , \tag{18}$$

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y , \tag{19}$$

$$Y' = \lambda Y^* + \epsilon_{y'} ; \tag{20}$$

where $d_y = 1$ if $Y' \geq M$, $d_y = 0$ if $Y' < M$, var $(Y') =$ var $(Y^*) = 1$, cov $(\epsilon_z,\epsilon_y) =$ cov $(\epsilon_y,\epsilon_{y'}) =$ cov $(\epsilon_z,\epsilon_y) = 0$, var $(\epsilon_z) = \sigma_{\epsilon_z}^2$, var $(\epsilon_y) = \sigma_{\epsilon_y}^2$, var $(\epsilon_{y'}) = \sigma_{\epsilon_{y'}}^2$, and all other notation is as described above. Equations (18), (19), and (20) can be expressed in reduced form for Z and Y' as follows:

$$Z = \gamma_0 + \gamma_1 X + \eta_z , \tag{21}$$

$$Y' = \theta_0 + \theta_1 X + \eta_{y'} ; \tag{22}$$

where

$$\gamma_0 = \alpha_0 + \alpha_2\beta_0 , \tag{23}$$

$$\gamma_1 = \alpha_1 + \alpha_2\beta_1 , \tag{24}$$

$$\theta_0 = \lambda\beta_0 , \tag{25}$$

$$\theta_1 = \lambda\beta_1 , \tag{26}$$

$$\begin{aligned} \text{cov}\,(\eta_z,\eta_{y'}) &= \text{cov}\,(\epsilon_{y'} + \lambda\epsilon_y, \epsilon_z + \alpha_2\epsilon_y) \\ &= \alpha_2\lambda\sigma_{\epsilon_y}^{\;2} , \end{aligned} \tag{27}$$

$$\text{var}\,(\eta_z) = \sigma_{\epsilon_z}^{\;2} + \alpha_2^{\;2}\sigma_{\epsilon_y}^{\;2} , \tag{28}$$

$$\text{var}\,(\eta_{y'}) = \sigma_{\epsilon_{y'}}^{\;2} + \lambda^2\sigma_{\epsilon_y}^{\;2} . \tag{29}$$

The reduced-form equations (21) and (22) can be estimated by OLS (for continuous Z) and either logit or probit analysis, respectively, providing estimates of γ_0, γ_1, θ_0, θ_1, cov $(\eta_z,\eta_{y'})$, var (η_z), and var $(\eta_{y'})$. These estimates, however, yield seven equations, (23)–(29), in eight unknown structural parameters: α_0, α_1, α_2, β_0, β_1, λ, σ_{ϵ_z}, and σ_{ϵ_y}. None of these parameters is identifiable. Note, however, that if $\lambda = 1$, model II reduces to model I and all parameters can be estimated as for the latter model.

Model II can, nonetheless, be identified in its own right with additional information. In particular, if there exists another exogenous variable, say W, that affects Y^* but not Z, model II can be identified. Alternatively, if there are two observed dichotomous variables that index Y^*, say d_{y1} and d_{y2}, instead of just one, then, too, the model can be identified. The following shows how an instrumental variable W can be used to identify model II. The use of multiple discrete indicators is outlined in the discussion for model IV below.

Modify model II by replacing (19) with

$$Y^* = \beta_0 + \beta_1 X + \beta_2 W + \epsilon_y \tag{19*}$$

but allowing (18) and (20) to remain as above. Then the reduced forms become

$$Z = \gamma_0 + \gamma_1 X + \gamma_2 W + \eta_z , \tag{21*}$$

$$Y' = \theta_0 + \theta_1 X + \theta_2 W + \eta_{y'} , \tag{22*}$$

where equations (23)–(29) hold as before and, in addition,

$$\theta_2 = \lambda\beta_2 , \tag{30}$$

$$\gamma_2 = \alpha_2\beta_2 . \tag{31}$$

Equations (23)–(31) now provide nine independent equations in the nine unknown structural parameters. These equations, along with the scaling assumptions on Y^* and Y', yield the following solutions for the structural parameters:

$$\lambda = \sqrt{\left[1 - \text{var}\,(\eta_{y'}) + \frac{\theta_2}{\gamma_2}\,\text{cov}\,(\eta_z,\eta_{y'})\right]} = k\ ;$$

$$\beta_0 = \frac{\theta_0}{k}\ ; \quad \beta_1 = \frac{\theta_1}{k}\ ; \quad \beta_2 = \frac{\theta_2}{k}\ ;$$

$$\alpha_0 = \gamma_0 - \frac{\gamma_2\theta_0}{\theta_2}\ ; \quad \alpha_1 = \gamma_1 - \frac{\gamma_2\theta_1}{\theta_2}\ ;$$

$$\alpha_2 = \frac{\gamma_2}{\theta_2}k\ ; \quad \sigma_{\epsilon_y} = \sqrt{\left[\frac{\theta_2}{\gamma_2}\,\text{cov}\,(\eta_z,\eta_{y'})\right]}/k\ ;$$

$$\sigma_{\epsilon_z} = \sqrt{\text{var}\,(\eta_z) - \frac{\gamma_2}{\theta_2}\,\text{cov}\,(\eta_z,\eta_{y'})}\ .$$

Thus the introduction of an instrumental variable W that affects Y^* but not Z allows the identification of all structural parameters. The single restriction that W does not affect Z compensates for the single additional parameter λ included in model II but not model I.

4. Model III

Recall that in model III the observed discrete response d_y is related deterministically to a latent continuous variable Y^* and has a direct impact on the endogenous variable Z (see fig. 5). That is, $Z = \alpha_0 + \alpha_1 X + \alpha_3 d_y + \epsilon_z$ and $Y^* = \beta_0 + \beta_1 X + \epsilon_y$, where var $(Y^*) = 1$, cov $(\epsilon_z,\epsilon_y) = 0$, and all notation is as defined above. Since all the independent variables are observed, model III can be estimated by the usual single-equation methods; that is, OLS for continuous Z or logit or probit analysis for discrete Z and logit or probit analysis for Y^*. Subject to the scale restrictions on Y^*, the model is identified directly and presents no special problems.

5. Model IV

Finally, consider model IV, which differs from model III only in that the relationship between the observed discrete variable and the latent continuous variable to which it corresponds is stochastic. This model, diagramed in figure 6*B*, is written as above:

$$Z = \alpha_0 + \alpha_1 X + \alpha_3 d_y + \epsilon_z , \tag{32}$$

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y , \tag{33}$$

$$Y' = \lambda Y^* + \epsilon_{y'} ; \tag{34}$$

where $d_y = 1$ if $Y' \geqslant M$, $d_y = 0$ if $Y' < M$, cov $(\epsilon_z, \epsilon_y) =$ cov $(\epsilon_y, \epsilon_{y'}) =$ cov $(\epsilon_z, \epsilon_{y'}) = 0$, var $(Y') =$ var $(Y^*) = 1$, and all other notation is as described above. Equation (32) can be estimated directly by OLS if Z is continuous or by logit or probit analysis if Z is discrete. Equations (33) and (34), however, are not identifiable. To see this, consider the reduced-form equation for Y', $Y' = \theta_0 + \theta_1 X + \eta_{y'}$, where $\theta_0 = \lambda\beta_0$, $\theta_1 = \lambda\beta_1$, and var $(\eta_{y'}) = \sigma_{\epsilon_{y'}}^2 + \lambda^2\sigma_{\epsilon_y}^2$. Since cov $(\eta_{y'}, \epsilon_z) = 0$, there are only three equations from which to estimate the four parameters λ, β_0, β_1, and $\sigma_{\epsilon_y}^2$. A key feature of this model is that, although the measurement parameter λ cannot be separated from the structural parameter β_1, the effect of X on Z via Y^* can nonetheless be calculated from the reduced-form parameter θ_1. This will be shown in the next section.

Unlike model II, which is also underidentified without additional information, model IV cannot be identified by resort to instrumental variables, that is, variables affecting Y^* but not Z. An alternative method that does permit identification, however, is to obtain additional categorical indicators of Y^*. Muthén (1979) shows that repeated observations on Y^* can be used to identify models similar to model IV. Consider an extension of model IV which retains (33) but replaces (32) and (34) with the three equations:

$$Z = \alpha_0 + \alpha_1 X + \alpha_3 d_{y_1} + \alpha_4 d_{y_2} + \epsilon_z , \tag{32*}$$

$$Y'_1 = \lambda_1 Y^* + \epsilon_{y'_1} , \tag{35}$$

$$Y'_2 = \lambda_2 Y^* + \epsilon_{y'_2} ; \tag{36}$$

where $d_{y_1} = 1$ if $Y'_1 \geqslant M_1$, $d_{y_1} = 0$ if $Y'_1 < M_1$, $d_{y_2} = 1$ if $Y'_2 \geqslant M_2$; $d_{y_2} = 0$ if $Y'_2 < M_2$; var $(Y'_1) =$ var $(Y'_2) = 1$; ϵ_z, $\epsilon_{y'_1}$, and $\epsilon_{y'_2}$ are uncorrelated; M_1 and M_2 are threshold parameters; and

$$\text{var}\,(\epsilon_{y'_1}) = \sigma_{\epsilon_{y'_1}}^2 \quad \text{and} \quad \text{var}\,(\epsilon_{y'_2}) = \sigma_{\epsilon_{y'_2}}^2 .$$

This model, which is diagramed in figure 7, contains two observed dichotomous indicators, d_{y_1} and d_{y_2}, of the underlying construct Y^*. As in the single-indicator case, Y'_1 and Y'_2 are related linearly but stochastically to Y^* but related deterministically to d_{y_1} and d_{y_2}, respectively. This model might be applied to an extension of the robbery victimization–gun ownership example discussed above. Let X denote personal background characteristics, Y^* denote neighborhood climate, d_{y_1} denote whether the individual has recently been robbed, d_{y_2} denote whether or not the individual knows someone else who has recently been robbed, and Z denote gun ownership. Model IV then models the effects of personal experiences with robbery on ownership.

This model can be identified as follows. Again the equation for Z can be estimated directly using a method appropriate to the way that Z is measured. To obtain the remaining parameters, solve (35) and (36) to obtain the reduced-form equations for Y'_1 and Y'_2: $Y'_1 = \theta_0 + \theta_1 X + \eta_{y'_1}$ and $Y'_2 = \psi_0 + \psi_1 X + \eta_{y'_2}$, where

$$\theta_0 = \beta_0\lambda_1 \; ; \quad \theta_1 = \beta_1\lambda_1 \; ; \quad \psi_0 = \beta_0\lambda_2 \; ; \quad \psi_1 = \beta_1\lambda_2 \; ;$$

$$\mathrm{cov}\,(\eta_{y'_1}, \eta_{y'_2}) = \lambda_1\lambda_2\sigma_{\epsilon_y}^2 \; ; \quad \mathrm{var}\,(\eta_{y'_1}) = \sigma_{\epsilon_{y'_1}}^2 + \lambda_1^2\sigma_{\epsilon_y}^2 \; ;$$

$$\mathrm{var}\,(\eta_{y'_2}) = \sigma_{\epsilon_{y'_2}}^2 + \lambda_2^2\sigma_{\epsilon}^2 \; .$$

These equations can be solved for the parameters of interest; that is,

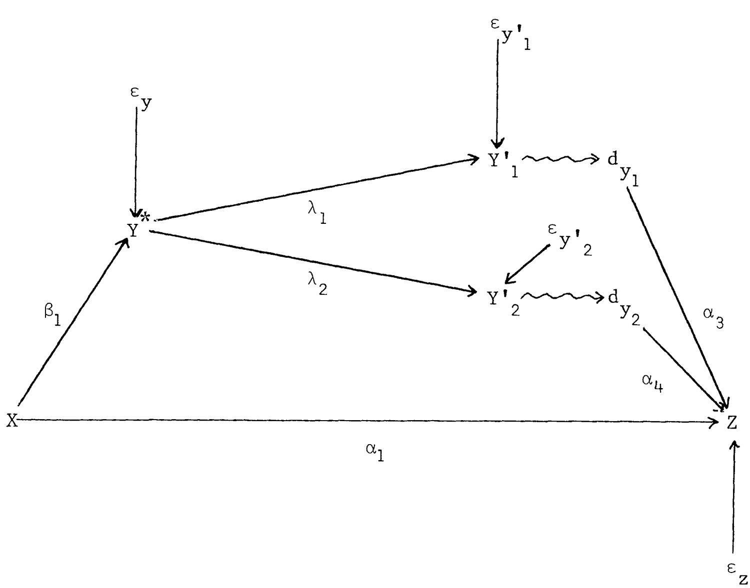

FIG. 7.—Threshold representation of model IV with multiple indicators

$$\lambda_1 = \sqrt{1 - \text{var}(\eta_{y'_1}) + \frac{\theta_1}{\psi_1}\text{cov}(\eta_{y'_1},\eta_{y'_2})} = k_1 ,$$

$$\lambda_2 = \sqrt{1 - \text{var}(\eta_{y'_2}) + \frac{\psi_1}{\theta_1}\text{cov}(\eta_{y'_1},\eta_{y'_2})} = k_2 ,$$

$$\beta_0 = \frac{\theta_0}{k_1}; \quad \beta_1 = \frac{\theta_1}{k_1}; \quad \sigma_{\epsilon_y}^{\,2} = \frac{\text{cov}(\eta_{y'_1},\eta_{y'_2})}{k_1 k_2} ,$$

$$\sigma_{\epsilon_{y'_1}}^{\,2} = 1 - k_1^{\,2} ; \quad \sigma_{\epsilon_{y'_2}}^{\,2} = 1 - k_2^{\,2} .$$

Thus multiple indicators of Y^* permit model IV to be identified.

The foregoing discussion has shown that the four basic recursive models vary considerably in their ease of identification. The alternative avenues to identification for these models are summarized in table 1.

Models I, II, III, and IV, of course, are not the only possible recursive models. An important type of model is one that combines the features of models I and III or those of models II and IV. That is, the effects of a discrete variable result from both the discrete variable, per se, and a latent continuous variable of which the discrete variable is an indicator. Indeed, such models are necessary to assess the relative importance of the features of models I and III or those of II and IV. It can be shown that a model that combines the features of models I and III can be identified under the same conditions as model I. That is, simple single-equation estimates will not suffice to identify the model, but using an estimate of the reduced-form error covariance (cov $[\eta_z,\epsilon_y]$), or an instrumental variable that either affects Y^* but not Z or has multiple indicators of Y^*, allows the model to be identified. Similarly, it can be shown that a model that combines the features of models II and IV can be identified under the same conditions as model IV. That is, multiple indicators of Y^* identify the model. These procedures are illustrated in Section IV below.

III. PATH ANALYSIS IN RECURSIVE MODELS WITH DISCRETE ENDOGENOUS VARIABLES

This section shows that the same principles of path analysis that have been applied to models with continuous dependent variables (Alwin and Hauser 1975; Duncan 1966) can be applied to the models considered in this article. That is, the total effect of a predetermined variable, say X, on an endogenous variable, say Z, can be apportioned into a component due to the indirect effect of X on Z through an intervening variable, say

Y, and a direct effect of X on Z net of Y. The "calculus of paths" is slightly more complex when Z or Y is discrete than when both are continuous because not all the relationships in the model are linear. Stolzenberg (1979), however, shows that such calculations are feasible for nonlinear models in general, and this section applies his methods to nonlinearities arising in models with discrete variables.

1. Continuous Endogenous Variables

Consider again the two-equation model for the ith observation: $Z = \alpha_0 + \alpha_1 X + \alpha_2 Y + \epsilon_z$, and $Y = \beta_0 + \beta_1 X + \epsilon_y$, where cov $(\epsilon_z, \epsilon_y) = 0$. If both Z and Y are continuous, the direct and indirect effects of X and Z are given by

$$\underset{\text{(total)}}{b_{zx}} = \underset{\text{(direct)}}{\alpha_1} + \underset{\text{(indirect)}}{\alpha_2 \beta_1}, \tag{37}$$

where b_{zx} is the zero-order (and reduced-form) regression of Z on X. This relationship holds whether Z and Y are observed continuous variables or latent variables with dichotomous indicators, say d_z and d_y. Thus (37) is an appropriate decomposition of the direct and indirect effects of X on Z for models I and II (see figs. 3 and 4), in which the effect of Y is hypothesized to be that of a latent continuous variable.

2. Continuous Intervening Variable; Discrete Z

The decomposition (37) is measured in units of the continuous variable Z, whether observed or unobserved. In any of models I through IV, however, when Z is a latent variable corresponding to a discrete variable d_z, the analyst may wish to calculate the effects of X on Z in terms of the discrete variable d_z rather than the continuous variable itself. It may, for example, be more meaningful to determine the effects of X on the probability that $d_z = 1$ than on the logit or probit transform of that probability. In this case, and in models III and IV, where the effect of Y is hypothesized to be that of an observed discrete variable d_y, (37) must be modified to take into account the nonlinear relationships between the discrete and continuous variables.

By elementary calculus, a dependent variable Z may be decomposed by its total differential $dZ = (\partial Z/\partial X)dX + (\partial Z/\partial Y)dY$, which implies the general formula

$$\frac{dZ}{dX} = \frac{\partial Z}{\partial X} + \frac{\partial Z}{\partial Y}\frac{dY}{dX}, \tag{38}$$

of which (37) is a special case (Stolzenberg 1979). Now let Z be a latent variable indicated by an observed dichotomous variable d_z and let Y be a continuous variable. This is again model I or II. It is possible to calculate the expected effects of X on d_z, that is, the effects of X on the expected values of d_z (given X).[14] The expected values of d_z are given by the equation $p(d_z = 1) = F(\alpha_0 + \alpha_1 X + \alpha_2 Y)$, where, as before, F is a general function that is typically the logistic or cumulative normal. Then the total effect of X on $p(d_z = 1)$ is, using (38), decomposed as

$$\begin{aligned} \frac{dp(d_z=1)}{dX} &= \alpha_1 f(\alpha_0+\alpha_1 X+\alpha_2 Y) + \alpha_2[f(\alpha_0+\alpha_1 X+\alpha_2 Y)]\frac{dY}{dX} \\ &= \alpha_1 f(\alpha_0+\alpha_1 X+\alpha_2 Y) + \alpha_2[f(\alpha_0+\alpha_1 X+\alpha_2 Y)]\beta_1 , \\ \text{(total)} \qquad & \qquad \text{(direct)} \qquad\qquad\qquad \text{(indirect)} \end{aligned} \tag{39}$$

where f is the derivative of F, that is, the probability density function corresponding to F (Hanushek and Jackson 1977). If F is the logistic function,

$$\begin{aligned} \frac{dp(d_z=1)}{dX} &= \frac{\alpha_1 \exp(\alpha_0+\alpha_1 X+\alpha_2 Y)}{[1+\exp(\alpha_0+\alpha_1 X+\alpha_2 Y)]^2} + \frac{\alpha_2\beta_1 \exp(\alpha_0+\alpha_1 X+\alpha_2 Y)}{[1+\exp(\alpha_0+\alpha_1 X+\alpha_2 Y)]^2} \\ &= \alpha_1 p_z(1-p_z) + \alpha_2\beta_1 p_z(1-p_z) , \end{aligned} \tag{40}$$

where $p_z = p(d_z = 1)$ is the probability that $d_z = 1$. If F is the cumulative normal distribution function,

$$\frac{dp(d_z = 1)}{dX} = \frac{\alpha_1 \exp(-t^2/2)}{\sqrt{2\pi}} + \frac{\alpha_2\beta_1 \exp(-t^2/2)}{\sqrt{2\pi}} , \tag{41}$$

where t is the standardized normal variable corresponding to $p(d_z = 1)$ under the probit transformation. The key feature of these decompositions

[14] The motivation for examining "expected effects" of independent variables on dichotomous dependent variables differs somewhat under the binomial trials and threshold views of discrete variables. In the binomial trials model, where each observation has an unobserved probability, say p, that it will take the value 1 on d_y, the expected effect can be interpreted as the expected change in p from observation to observation for a change in the value of X. In the threshold model, in contrast, changes in X change d_y only when Y^* is at its threshold L; i.e., $d(d_y)/dY^* = 0$ if $Y^* \neq L$, $d(d_y)/dY^* = \pm\infty$ if $Y^* = L$. In this case the expected effect can be interpreted as the effect an observer might expect for a group of observations with a common value on X; i.e., the effect is the average for the population of observations with the same values on X. If the observations refer to individuals, then, in the binomial trials model, the estimated effects are those that can be expected for both the individual and the investigator. In the threshold model, the estimated effects are expected only by the investigator who examines a group of individuals with a common value on X.

is that they depend on the observations being considered; that is, the effects of X on d_z depend on the level of X. This is, of course, exactly what the nonlinear hypothesis of the probit or logit model requires. In practice, therefore, one should calculate the effects of X for a range of values of X or p to chart the nonlinear relationship. Among other values, perhaps the most useful summary calculation is to choose the sample proportion of cases for which $d_z = 1$. This proportion can be substituted for p_z in (40) or its normal probability transform t can be substituted in (41). Note, however, that the *relative* importance of the direct and indirect effects is invariant under the choice of X or of p because $f(\alpha_0 + \alpha_1 X + \alpha_2 Y)$ is simply a constant of proportionality in (39). Moreover, for models I and II, the relative importance will be the same whether the effects of X are decomposed in the scale of the latent continuous variable Z as in (37) or in the scale of $p(d_z = 1)$ as in (39).

3. Discrete Intervening Variable

The preceding discussion assumed that the intervening variable Y affected Z as a continuous variable, either latent or observed. Now consider the decomposition for models III and IV (figs. 5 and 6) where the discrete variable d_y affects Z directly. The model is $Z = \alpha_0 + \alpha_1 X + \alpha_3 d_y + \epsilon_z$, and $Y = \beta_0 + \beta_1 X + \epsilon_y$, where $\text{cov}(\epsilon_z, \epsilon_y) = 0$.

Model III.—For decomposition it is necessary to take account of the nonlinear relationship between d_y and X. The expected value of d_y is $p(d_y = 1) = F(\beta_0 + \beta_1 X)$, where F is typically the logistic or cumulative normal function, and thus the expected effect of X on d_y is given by $dp(d_y = 1)/dX = \beta_1 f(\beta_0 + \beta_1 X)$, where f is the probability density function, either extreme value or normal, corresponding to F (Hanushek and Jackson 1977). Now if Z is continuous, the effect of X on Z[15] is

$$\begin{aligned} \frac{dZ}{dX} &= \alpha_1 + \frac{\alpha_3 dp(d_y = 1)}{dX} \\ &= \alpha_1 + \alpha_3 \beta_1 f(\beta_0 + \beta_1 X)\ . \\ &\text{(total)} \quad \text{(direct)} \quad \text{(indirect)} \end{aligned} \tag{42}$$

If F is logistic, $dZ/dX = \alpha_1 + \alpha_3 \beta_1 p_y(1 - p_y)$, where $p_y = p(d_y = 1)$ and $p_y = \exp(\beta_0 + \beta_1 X)/[1 + \exp(\beta_0 + \beta_1 X)]$. If F is the cumulative normal function, $dZ/dX = \alpha_1 + \alpha_3 \beta_1 [\exp(-t^2/2)/\sqrt{2\pi}]$. If Z is a discrete

[15] More precisely, this formula gives the *expected* effect of X on Z, i.e., $dE(Z)/dX$, inasmuch as the indirect effect of X on Z includes the effect of X on $p(d_y = 1)$. For expository convenience, this mathematical nuance is ignored in the formulas presented in this section.

variable and the effects of X on the expected values of d_z are required, the decomposition is

$$\underset{\text{(total)}}{\frac{dp(d_z = 1)}{dX}} = \underset{\text{(direct)}}{\alpha_1 f(\alpha_0 + \alpha_1 X + \alpha_3 d_y)} + \underset{\text{(indirect)}}{\alpha_3 f(\alpha_0 + \alpha_1 X + \alpha_3 d_y)\beta_1 f(\beta_0 + \beta_1 X)} \,. \tag{43}$$

As before, f is the extreme value density function under a logit model or the standard normal density function under the probit model. In the former case, for example, the decomposition is $dp(d_z = 1)/dX = \alpha_1 p_z(1 - p_z) + \alpha_3 p_z(1 - p_z)\beta_1 p_y(1 - p_y)$, where specific values of $p_z = p(d_z = 1)$ and $p_y = p(d_y = 1)$, say the sample proportions for which $d_z = 1$ and for which $d_y = 1$, must be chosen.

Model IV.—In addition to the structural equations given above, model IV includes the measurement equation $Y' = \lambda Y + \epsilon_{y'}$, indicating that the link between Y and d_z is stochastic as well as nonlinear. Under this model, $dp(d_y = 1)/dX = [dp(d_y = 1)/dY\,]\, dY/dX = \beta_1\lambda f(\beta_0 + \beta_1 X)$. If Z is continuous, the decomposition of direct and indirect effects of X on Z is

$$\underset{\text{(total)}}{\frac{dZ}{dX}} = \underset{\text{(direct)}}{\alpha_1} + \underset{\text{(indirect)}}{\alpha_3\beta_1\lambda f(\beta_0 + \beta_1 X)} \,.$$

If the effects of X on d_z are required, the decomposition is

$$\underset{\text{(total)}}{\frac{dp(d_z = 1)}{dX}} = \underset{\text{(direct)}}{\alpha_1 f(\alpha_0 + \alpha_1 X + \alpha_3 d_y)} + \underset{\text{(indirect)}}{\alpha_3[f(\alpha_0 + \alpha_1 X + \alpha_3 d_y)]\beta_1\lambda[f(\beta_0 + \beta_1 X)]} \,.$$

Here and in (42) and (43) the relative sizes of the direct and indirect components are unchanged whether one calculates the effects on Z or on $p(d_z = 1)$. Note also that for model IV it is unnecessary to identify the measurement parameter λ to calculate the path decomposition. It suffices to be able to identify $\beta_1\lambda$ since only this product is used in the calculations. Thus the decomposition for model IV is feasible under broader circumstances than those in which the model itself can be identified.

IV. EMPIRICAL EXAMPLES

This section illustrates the models and methods presented above with two numerical examples. The first applies models I and III to the effects of social background and parental factors on college plans of high school seniors. The second applies models II and IV to the effects of state Fair-Employment-Practices Legislation and other social factors on the relative earnings of black and white workers.

1. Social Background, Parental Encouragement, and College Plans

Using a sample of high school seniors, Sewell and Shah (1968) and Fienberg (1980) examine the effects of social background factors, including sex, socioeconomic status, and intelligence, on the extent to which their parents encourage them to attend college, and the effects of both social background and parental encouragement on whether the seniors themselves plan to attend college. The researchers apply a simple recursive model with two endogenous variables: parental encouragement and college plans. Encouragement is interpreted as intervening causally between social background and plans. Both endogenous variables are measured discretely. Whether or not the senior plans to attend college is coded yes or no. Parental encouragement to attend college is coded as high or low. Previous analyses of these data have examined the effect on college plans of a dummy variable for parental encouragement, with the social background variables controlled, and the effects of the background variables on parental encouragement. This two-equation model, therefore, is an instance of either model III or model IV discussed above. Parental encouragement, however, may be better modeled as resulting from variation on an unmeasured continuous variable indicating that degree of encouragement is a continuum from none to a great deal, than as a simple binary difference between those coded as experiencing "high" or "low" encouragement. In this case, the recursive model of choice is either model I or model II discussed above since these models allow the effects of discrete intervening variables to result from continuous latent variables. These models do not necessarily fit the data equally well. Nor do they have the same implications for the degree to which parental encouragement mediates the effect of social background on college plans.

Data.—The data are a cross-classification of the characteristics of 10,318 Wisconsin high school seniors in 1957. The characteristics and their measurement are as follows: socioeconomic status (S) (high, upper-middle, lower-middle, low), intelligence (Q) as measured by the Hammon-Nelson Test of Mental Ability (high, upper-middle, lower-middle, low), sex (X) (female, male), parental encouragement (P) (high, low), and college plans

(C) (yes, no). The data are reported by Sewell and Shah (1968) and by Fienberg (1980, p. 130).[16] In the models estimated here, both IQ and socioeconomic status are coded as continuous variables scaled to have means of zero and variances of unity.[17]

Models.—As noted above, it is of interest to contrast alternative roles of parental encouragement in recursive models of background effects on college plans. To carry this out, models I and III and a model that combines the effects of models I and III are estimated. Models I and III are nested within the latter model, and thus their relative fits to the data can be derived from the three models. Models I and III rather than models II and IV are considered here because the former are identifiable without additional assumptions, whereas the latter are not. In modeling parental encouragement, moreover, models I and III are substantively plausible. That is, the continuous underlying variable for parental encouragement may be reasonably assumed to be classified perfectly by the observed dichotomous variable. That is, there are no individuals with high scores on the unmeasured encouragement measure who are coded "low" on the observed variable or vice versa. Measurement error is, of course, always possible, but there is no other substantive reason for modeling the link between the continuous and discrete variables as stochastic.

For models I, III, and their combination, the effects of social background can be modeled as $P^* = \beta_0 + \beta_1 X + \beta_2 Q + \beta_3 S + \epsilon_p$, where ϵ_p is a normally distributed stochastic disturbance, P^* is an unmeasured continuous variable, and P = "high" if $P^* \geqslant L$, P = "low" if $P^* < L$, where L is an unobserved threshold. For model I the effects of background and parental encouragement can be written $C^* = \alpha_0 + \alpha_1 X + \alpha_2 Q + \alpha_3 S + \alpha_4 P^* + \epsilon_c$, where ϵ_c is a normally distributed disturbance, C^* is an unmeasured continuous variable, and C = "yes" if $C^* \geqslant N$, C = "no" if $C^* < N$, where N is an unobserved threshold. For model III the effect of P on C is discrete, and the model can be written $C^* = \alpha_0 + \alpha_1 X + \alpha_2 Q + \alpha_3 S + \alpha_5 P + \epsilon_c$. Finally, the combined model is $C^* = \alpha_0 +$

[16] The data are also available from us on request.

[17] Fienberg's (1980) reported results as well as additional analyses of the data show that some of the effects of socioeconomic status and intelligence in the model are nonlinear. Fienberg also reports significant three-way interactions among pairs of the background variables and the endogenous variables. The models reported here assume both linearity of the background effects and additivity of all effects in the model. These simplifications clarify the exposition and yield substantively plausible results. The models presented here do not fit the data as well as the more complex models estimated by Fienberg, and the differences in fit between models I and III reported below *may* result in part from failure to fit higher-order interactions to the data. It is nonetheless straightforward to incorporate nonlinearities and higher-order interactions into the models presented here. However, in practice, models with higher-order interactions and models with latent continuous variables may be empirically indistinguishable without prior conceptions of which specification is more plausible.

$\alpha_1 X + \alpha_2 Q + \alpha_3 S + \alpha_4 P^* + \alpha_5 P + \epsilon_c$. For all models, ϵ_c and ϵ_p are assumed to be uncorrelated.

Identification and estimation.—Models I, III, and their combination are identifiable by the arguments presented above. Model III can be identified directly from the parameters of single-equation estimates of the determinants of parental encouragement and college plans. Model I and the model that combines the effects of models I and III can be identified and estimated from the parameters of the equation for parental encouragement, a reduced-form equation for college plans, and the disturbance covariance for the two reduced-form equations. Given the normality assumption of the errors in the structural equations, the models were estimated by maximum likelihood probit analysis (see Appendix).[18]

Results.—Table 2 presents the log likelihood statistics for four models estimated by maximum likelihood probit analysis. From these models can be derived the structural parameters for the models of interest. The four include a model in which (1) there is no effect of parental encouragement

TABLE 2

LOG LIKELIHOODS AND LIKELIHOOD RATIO χ^2 TESTS FOR ALTERNATIVE MODELS OF EFFECTS OF SOCIAL BACKGROUND ON PARENTAL ENCOURAGEMENT AND COLLEGE PLANS

A. LOG LIKELIHOODS

Model*	Log Likelihood
A. No effect of parental encouragement on college plans.	−10,903 21
B. Effect of observed dummy variable for high parental encouragement on college plans (model III)	−10,078.14
C. Effect of unobserved continuous variable for parental encouragement on college plans (model I)	−10,069.75
D. Effect of both unobserved continuous and observed dummy variables on college plans (models I and III).	−10,069.06

B. LIKELIHOOD RATIOS

Model Comparison	Likelihood Ratio χ^2†	Degrees of Freedom	P
B vs. A	1,650.1	1	<.001
C vs. A	1,666.9	1	<.001
D vs. B	18.2	1	<.001
D vs. C	1.4	1	$.2<P<.3$

* All models include effect of sex, IQ, and parental socioeconomic status on both parental encouragement and college plans See text for discussion of data and measurement

† Likelihood ratio χ^2 statistics are computed as −2 times the differences of the appropriate log likelihoods

[18] A computer program for estimating two reduced-form equations with binary dependent variables and correlated disturbances by maximum likelihood probit analysis is available from us.

on college plans, (2) the dummy variable for encouragement affects college plans but the reduced-form disturbance covariance is assumed to be zero (model III), (3) the dummy variable for encouragement does not affect college plans but the disturbance covariance is nonzero (model I), and (4) the disturbance covariance is nonzero and there is a dummy variable effect of encouragement on college plans (models I and III combined).

The log likelihoods in table 2 indicate the relative fits of alternative models for the parental encouragement effect on college plans. When the log likelihood of a model is subtracted from the log likelihood for a model within which the first model is nested, -2 times the difference is distributed as χ^2 with degrees of freedom (*df*) equaling the difference in numbers of parameters between the two models (e.g., Hogg and Craig 1970). The lower panel of table 2 contrasts several pairs of models. The first two lines show the improvement in fit of the model when the two types of parental encouragement effects on college plans are added separately to a model that includes only background effects on college plans and on parental encouragement. The dummy and continuous latent variable effects reduce χ^2 by 1,650 and 1,667, respectively, both highly significant improvements in fit. To see whether both types of effects are necessary, a model that combines the effects of models I and III is contrasted with the models that include either the dummy variable or the latent continuous variable for parental encouragement. These contrasts, shown in the last two lines of table 2, show that whereas the elimination of the continuous latent variable from the model significantly weakens the fit of the model ($\chi^2 = 18.2$ with 1 *df*), the exclusion of the dummy variable has no significant impact on the fit of the model ($\chi^2 = 1.4$ with 1 *df*). Thus, relative to the combined model, model I, which includes the effect of a latent continuous variable for encouragement and excludes the dummy variable, cannot be rejected, but model III, which includes the encouragement effect as a dummy variable and excludes the latent variable, is rejected.

Table 3 presents the reduced-form probit coefficients for the determinants of parental encouragement and college plans under alternative models. These coefficients are expressed in raw form under the assumption that the disturbance variances of the equations are unity. As indicated by the asymptotic normal statistics, both the effects of the dummy variable for parental encouragement in model III and the disturbance correlation, from which is derived the effect of the latent variable for encouragement in model I, are highly significant (see the third and fourth columns). When both the error correlation and the dummy variables are included in the model, as shown in the last column, the error correlation remains highly significant, but the dummy variable coefficient is negative and insignificant. This is consistent with the goodness-of-fit results reported in table

TABLE 3

REDUCED-FORM MAXIMUM LIKELIHOOD PROBIT PARAMETER ESTIMATES FOR EFFECTS OF BACKGROUND FACTORS ON PARENTAL ENCOURAGEMENT AND COLLEGE PLANS

	Dependent Variable				
		College Plans			
		Corresponding Structural Model			
Independent Variable	Parental Encouragement*	No Parental Encouragement Effect	III	I	I and III
Sex (female)	−.359	−.255	−.121	−.254	−.283
	(−13.3)	(−9.0)	(−3.9)	(9.0)	(7.7)
IQ	.334	.452	.370	.450	.447
	(24.1)	(30.6)	(23.0)	(30.8)	(26.6)
SES	.538	.489	.302	.485	.514
	(37.3)	(32.5)	(17.8)	(32.5)	(19.6)
Parental encouragement (observed dummy variable)	...	...	1.372	...	−.383
			(38.1)		(−1.1)
ρ (correlation of errors in reduced-form equations)	...	...	...	.695	.846
				(57.2)	(6.2)
Constant	.250	.441	1.360	.438	.127
	(12.1)	(21.6)	(39.2)	(21.6)	(.41)

NOTE —Coefficients are unscaled raw probit estimates which need to be rescaled if error variances are not assumed to be unity Numbers in parentheses are asymptotic normal statistics (Z-scores)

* Parental encouragement coefficients are invariant under choice of model

2. These results suggest that parental encouragement is better modeled as a latent continuous variable in its effects on college plans than as the discrete variable that is directly measured.

Table 4 presents the structural parameters that are derived from the coefficients in table 3 through the identification methods presented in Section II. The first column presents the parental encouragement equation, and the second, third, and fourth columns present the coefficients for the reduced form and models III and I, respectively.[19] To obtain the coefficients for the parental encouragement equation in table 4, the raw coefficients are rescaled under the assumption that the latent variable for encouragement has variance of unity. This requires an estimate of the error variance which is, using the results of Section II.1,

$$\sigma_{\epsilon_p}^2 = 1/[b_1^2 \text{ var } (X) + b_2^2 \text{ var } (Q) + b_3^2 \text{ var } (S) + 2b_1b_2 \text{ cov } (X,Q) + 2b_1b_3 \text{ cov } (X,S) + 2b_2b_3 \text{ cov } (Q,S) + 1] ,$$

TABLE 4

STRUCTURAL COEFFICIENTS FOR EFFECTS OF BACKGROUND FACTORS ON PARENTAL ENCOURAGEMENT AND COLLEGE PLANS UNDER MODELS I AND III

	DEPENDENT VARIABLE*			
		College Plans		
			Model	
INDEPENDENT VARIABLE	PARENTAL ENCOURAGEMENT	Reduced Form	III	I
Sex (female)	−.290	−.203	−.083	−.004
	(−13.3)	(−9.0)	(−3.9)	(−0.2)
IQ	.269	.360	.254	.175
	(24.1)	(30.6)	(23.0)	(17.6)
SES	.434	.389	.208	.091
	(37.3)	(32.5)	(17.8)	(8.2)
Parental encouragement (observed dummy)	...	...	.942	...
			(38.1)	
Parental encouragement (unobserved continuous)	...	...	...	.686
				(57.2)
Constant	.202	.351	.934	.251
	(12.1)	(21.6)	(39.2)	(17.7)
Error variance	.650	.633	.472	.327

NOTE.—Numbers in parentheses are asymptotic normal statistics (Z-scores)

* Calculations assume that unmeasured continuous variables for parental encouragement and college plans have variances of unity

[19] Structural coefficients for the model combining the effects of models I and III are similar to those for model I.

where β_1, β_2, and β_3 are the raw coefficients reported in the first column of table 3. The variances and covariances of the observed independent variables are as follows:

	X	Q	S	P
X	.250	−.006	−.008	−.031
Q		1.000	.278	.162
S			1.000	.214
P				.250

Thus

$$\sigma_{\epsilon_p}^2 = 1/[(-.359)^2(.250) + (.334)^2 + (.538)^2 + 2(-.359)(.334)(-.006)$$
$$+ 2(-.359)(.538)(-.008) + 2(.334)(.538)(.278) + 1]$$
$$= .650.$$

With this estimate in hand, the rescaled coefficients can be obtained directly. For example, the rescaled coefficient for sex is $(-.359)(\sqrt{.650}$ $= -.290$. The coefficients for IQ, socioeconomic status, and the constant are similarly calculated as the products of σ_{ϵ_p} and the unscaled coefficients in table 3. The structural coefficients and error variances for the reduced-form and model III college-plans equations are derived by the same method.

The derivation of the structural parameters for model I, shown in the last column of table 4, follows the method outlined in Section II.2. The reduced-form error covariance for the parental encouragement and college-plans equation is

$$\rho\sqrt{\sigma_{\eta_c}^2\sigma_{\eta_p}^2}\ ,$$

where ρ is the disturbance correlation and $\sigma_{\epsilon_p}^2$ and $\sigma_{\eta_c}^2$ are the error variances of the reduced-form equations for parental encouragement and college plans, respectively. Thus cov $(\eta_c, \epsilon_p) = .695\sqrt{(.650)(.633)} = .446$. Then the structural coefficient for the effect of the latent variable for parental encouragement on college plans is $.446/.650 = .686$ reported in the final column of table 4. The remaining structural coefficients for the social background effects in model I are calculated from the reduced-form equations using the formulas in Section II.2. The coefficient for sex, for example, is $(-.203) - (-.290)(.686) = -.004$. The IQ, SES, and constant coefficients are calculated similarly. Finally, the error variance for the college-plans equation under model I is $\sigma_{\epsilon_c}^2 = \sigma_{\eta_c}^2 - \alpha_3^2\sigma_{\epsilon_p}^2 =$ $.633 - (.686)^2(.650) = .327$.

The structural coefficients for the college-plans equation differ substantially between models I and III. The coefficients of the social background factors are smaller in model I, indicating weaker net effects than in model III. The coefficients for the two parental encouragement variables are not comparable inasmuch as they are measured in different scales. In model III the coefficient indicates that, for a latent variable indexing intention to attend college, there is a .942 standard deviation (SD) difference between individuals coded as high and those coded as low on parental encouragement. Under model I, the coefficient indicates that a 1 SD change in parental encouragement yields a .686 SD change in the latent variable for college plans. Finally, the error variance for model I is approximately 30% smaller than for model III, indicating that the model in which parental encouragement is treated as continuous is more successful in explaining college plans.

Direct and indirect effects of social background.—Models I and III have different implications for the allocation of the effects of social background on college plans into direct effects and indirect effects through parental encouragement. The decomposition of direct and indirect effects under the two models is shown in table 5. The top panel of the table shows the effects in SD units of the latent continuous variable for college plans. The bottom panel expresses the direct and indirect components as proportions of the total (reduced-form) effects of social background on

TABLE 5

DECOMPOSITION OF EFFECTS OF SEX, IQ, AND SOCIOECONOMIC STATUS ON COLLEGE PLANS UNDER ALTERNATIVE MODELS

	MODEL I			MODEL III		
VARIABLE	Total	Direct	Indirect via Parental Encouragement	Total	Direct	Indirect via Parental Encouragement
Absolute effects:						
Sex	−.203	−.004	−.199	−.192	−.083	−.109
IQ	.360	.175	.185	.345	.254	.100
SES	.389	.091	.298	.371	.208	.163
Relative effects:						
Sex	1.000	.020	.980	1.000	.432	.568
IQ	1.000	.486	.514	1.000	.726	.290
SES	1.000	.234	.766	1.000	.561	.439

NOTE —College plans are measured in probit transformed scale of probabilities of planning to attend college, that is, $\phi^{-2}(p)$, where ϕ is the cumulative normal function and p is the probability of planning to attend college To obtain decomposition of the actual probability of attending college, multiply each entry by $(1/\sqrt{2\pi})\exp(-z^2/2)$, where

$$p = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(\frac{-u^2}{2}\right) du$$

for a selected value of p (Hanushek and Jackson 1977, p 189) At $p =$ 327, the overall sample proportion planning to attend college, $z = -$ 448 and $(1/\sqrt{2\pi})\exp(-z^2/2) =$ 361 Relative effects, however, are invariant under this transformation college plans. For model I, the absolute effects are calculated using the basic methods of path analysis. The direct effects are simply the structural coefficients for the background variables reported in the last column of table 4. The indirect effects are the products of the social background coefficients in the parental encouragement equation and the structural parameter for encouragement in the model I college-plans equation. For model III, the direct effects are again the structural coefficients for the social background factors obtained under that model and reported in the third column of table 4. The indirect effects are the products of three factors: the coefficient for the independent variable in the encouragement equation, the coefficient for the dummy variable for encouragement in the college-plans equation, and a factor $\exp(-z^2/2)\sqrt{2\pi}$, where

$$p = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(\frac{-u^2}{2}\right) du$$

for a selected value of p. This last factor is simply the ordinate of the normal curve corresponding to the "area" under the curve measured by p. Taking p to be the sample proportion with "high" parental encouragement, $p = .5188$, $z = .0471$, and $\exp(-z^2/2)/\sqrt{2\pi} = .3985$. Then, for example, the indirect effect of sex on college plans is $(-.290)(.942)(.3985) = -.1089$.

The table of relative effects shows that under model I, a much larger proportion of the social background influences on college plans occurs through differential parental encouragement than under model III. For example, according to model I, three-fourths of the effect of parental socioeconomic status on college plans can be attributed to the effects of SES on parental encouragement and the effect of the latter on college plans. Under model III, however, less than one-half of the SES effect on college plans is due to the intervening mechanisms of parental encouragement. These results reflect the more powerful influence of the continuous latent variable for encouragement relative to that of the observed dichotomous variable.

2. Fair-Employment-Practices Legislation and Relative Wages

Landes (1968) and Heckman (1976) examine the possible effects of state-level Fair-Employment-Practices Legislation (FEPL) on the relative wages of blacks and whites. The correlation between the existence of FEPL and the level of black wages relative to that of whites is positive, but such a correlation may arise through several mechanisms. The legislation itself may have an active effect by deterring employers who would otherwise discriminate against blacks in wages. Alternatively, the law itself may have no effect; it may be symptomatic of other social and historical factors that lead states not only to pass progressive laws but also to treat the races equally in the labor market. Under this Durkheimian view of the law, the accumulated sentiments in a state, not their legal manifestations, affect the relative wages of blacks and whites. For the purposes of this discussion, this factor will be termed "progressive sentiment." A third hypothesis is that neither the law nor progressive social conditions but other characteristics of states correlated with the existence of FEPL explain the positive zero-order correlations between the existence of FEPL and the relative standing of blacks.

These competing views can be represented in models II and IV. Under model IV, FEPL affects relative wages directly. Under model II, an unmeasured variable, of which FEPL is an indicator, affects relative wages, but FEPL itself does not. This latent variable can be interpreted as degree of progressive sentiment. In these models, the relationship between sentiment and FEPL is stochastic; that is, despite the positive association between FEPL and progressive sentiment, some progressive states may lack FEPL and, conversely, states where sentiments favoring discrimination are widespread may have FEPL. Under both models II and IV, progressive sentiment intervenes causally between other social factors and the relative wage levels of blacks and whites. Model IV cannot be identified with only a single indicator of the unmeasured variable. Moreover, progressive sentiment is better modeled as having several indicators. Thus a second indicator of progressive sentiment, based on state voting behavior, is included in the models discussed below.

Data and models.—The data for this example derive from published sources for the 48 continental United States during the 1950s and 1960s.[20] The relative wages of blacks and whites are measured as the ratio of wages of black males to those of white males in 1959 (B). Progressive sentiment (P^*) is indicated by two binary variables: (1) whether a state had a Fair-Employment-Practices Law before 1959 (F), and (2) whether a state gave less than 10% of its vote to George Wallace in the 1968 presidential election (W).[21] Progressive sentiment is assumed to arise from past and contemporary social conditions. To reflect this, two independent variables are assumed to affect sentiment, the percentage of the civilian labor force that was unionized in 1953 (N) and the percentage of the white work force with more than 12 years of schooling in 1959 (E). Both vari-

[20] For further documentation of the variables discussed below, see Landes (1968). The data are available from us.

[21] The 10% cutoff point on the Wallace vote is arbitrary. Two dichotomous indicators of progressive sentiment are used in this example to follow the general form of model IV as discussed in Section II. A better formulation, which could be estimated by simple extensions of the methods discussed here, would treat the Wallace vote as a continuous indicator of progressive sentiment.

ables are assumed to affect progressive sentiment positively and to increase the likelihood of FEPL and a low Wallace vote. As discussed thus far, the model is $P^* = \beta_0 + \beta_1 N + \beta_2 E + \epsilon_p$, where ϵ_p is a normally distributed disturbance and var $(P^*) = 1$. In addition, $F' = \lambda_F P^* + \epsilon_F$, and $W' = \lambda_W P^* + \epsilon_W$, where var (F') = var $(W') = 1$; ϵ_p, ϵ_F, and ϵ_W are mutually uncorrelated; $F = 1$ if $F' \geqslant M_1$; $W = 1$ if $W' \geqslant M_2$; $F = 0$ if $F' < M_1$; $W = 0$ if $W' < M_2$; and M_1 and M_2 are unobserved thresholds. This part of the model is similar to the relationships among X, Y^*, Y_1, Y_2, d_{y_1}, and d_{y_2} in figure 7.

In addition to FEPL and progressive sentiment, several other variables are assumed to affect relative wages. These include the ratio of nonwhite to white males in the civilian labor force in 1959 (L); the ratio of the percentage of the nonwhite male population living in urban areas to the corresponding percentage for white males in 1959 (R); the percentage of the total male population living in urban areas in 1959 (U); and the ratio of nonwhite male to white male mean years of schooling completed in 1959 (S). Then a general equation for the relative wages can be written $B = \alpha_0 + \alpha_1 L + \alpha_2 R + \alpha_3 U + \alpha_4 S + \alpha_5 F + \alpha_6 P^* + \epsilon_B$, where ϵ_B is a normally distributed disturbance that is uncorrelated with ϵ_p, ϵ_F, and ϵ_B. Under model II, $\alpha_5 = 0$, whereas under model IV, $\alpha_6 = 0$. If neither FEPL nor the sentiment that it indicates affects relative wages, then $\alpha_5 = \alpha_6 = 0$.[22]

Identification and estimation.—These models are identified using the methods discussed in Section II.5 for model IV. The measurement and structural equations for progressive sentiment can be identified separately from the relative wage equation. The models estimated here are overidentified inasmuch as each endogenous variable is affected by several exogenous variables that do not affect each other. Because no computer software that imposes the overidentifying restrictions is available, the estimates reported below are not unique but instead depend on the equations used in the solution. Moreover, since the sentiment and wage equations are estimated separately, no global goodness-of-fit statistics for the models are available. The R^2s of the alternative relative wage equations are compared in the discussion below, but they do not provide valid tests of the fits of the models.

Results.—Table 6 contains the reduced-form estimates for the equations predicting the two indicators of progressive sentiment. These estimates

[22] The rationale for these independent variables is discussed in Landes (1968). Heckman (1976) investigates the possible effects of FEPL and of progressive sentiment using somewhat different models from those investigated here His models treat FEPL as related deterministically to the latent variable for sentiment; in this they are similar to models I and III. Heckman also explores, however, the possible simultaneity of legislation and relative wages, an issue not considered here.

are scaled such that their corresponding latent variables F' and W' have variances of unity. From these parameters the measurement and structural parameters for progressive sentiment are obtained. Thus, from the formulas given in Section II.5, and arbitrarily using the coefficient on schooling, measurement parameters are given by

$$\lambda_F = \sqrt{1 - .572 + \frac{(.140)}{(.141)}(.620)(.756)(.856)} = .909 ,$$

$$\lambda_W = \sqrt{1 - .731 + \frac{(.141)}{(.140)}(.620)(.756)(.856)} = .820 ,$$

and

$$\sigma_{\epsilon_F}{}^2 = 1 - (.909)^2 = .173 ; \quad \sigma_{\epsilon_W}{}^2 = 1 - (.820)^2 = .328 ,$$

$$\sigma_{\epsilon_p}{}^2 = \frac{(.620)(.756)(.856)}{(.909)(.820)} = .538 .$$

Given the measurement parameters, consistent estimates of the structural equations for P^* can be obtained by arbitrarily focusing on the parameter for FEPL: $\beta_0 = -5.897/.909 = -6.487$; $\beta_1 = .109/.909 = .120$; $\beta_2 = .140/.909 = .154$. These are reported in the first column of table 7.

TABLE 6

REDUCED-FORM MAXIMUM LIKELIHOOD PROBIT PARAMETER ESTIMATES FOR EFFECT OF LEVELS OF SCHOOLING AND UNIONIZATION ON INDICATORS OF PROGRESSIVE SENTIMENT

	DEPENDENT VARIABLES	
INDEPENDENT VARIABLES*	FEPL (*F*)	Wallace Vote (*W*)
Unionization (*N*)	.101	.056
	(3.3)	(2 1)
Schooling (*E*)	.140	.141
	(1.3)	(1.3)
Constant	−5.897	−3.716
	(−2.5)	(−1.8)
Error variance	.572	.731
Disturbance correlation	.620	
	(2.4)	
Log likelihood	−48.39	
N	48	

NOTE —Coefficients are scaled under assumption that latent variables for FEPL and Wallace vote have variance of unity FEPL denotes whether or not a state had a Fair-Employment-Practices Law before 1959 Wallace vote denotes whether or not a state gave less than 10% of its popular vote to George Wallace in the 1968 presidential election Numbers in parentheses are asymptotic normal statistics (Z-scores)

* For discussion of independent variables, see text

The equations for relative wages in which neither F nor P^* has an effect and in which only F has an effect are estimated directly by OLS. The results are reported in the second and fourth columns of table 7. The equations for relative wages that include an effect for progressive sentiment and that include both sentiment and FEPL effects are estimated as follows. First, estimated values of P^* are obtained using the structural parameters reported in the first column of table 7, that is, $\hat{P}^* = -6.487 + .120N + .154E$ for each observation. Then the $\hat{P}^*$s are entered into the relative wage equations in place of the unobserved P^*. These equations are estimated by OLS and yield estimates for model II and the combined effects for models II and IV which are reported in the third and fifth columns of table 7.[23]

TABLE 7

STRUCTURAL COEFFICIENTS FOR EFFECTS OF SOCIAL FACTORS ON PROGRESSIVE SENTIMENT AND RELATIVE WAGES UNDER MODELS II AND IV

	Dependent Variables				
		Relative Wages (B)			
		Model			
Independent Variables*	Progressive Sentiment	No Effect of Law	II	IV	II and IV Combined
Unionization (N)	.120	...	...	...	...
Schooling (E)	.154	...	...	...	...
Relative labor force participation (L)	...	−.651	−.563	−.688	−.608
		(.126)	(.115)	(.115)	(.110)
Relative schooling (S)	...	.115	.081	−.010	−.009
		(.146)	(.130)	(.139)	(.128)
Relative urbanization (R)	...	.099	.095	.108	.103
		(.032)	(.029)	(.030)	(.027)
Urbanization (U)	...	−.033	−.134	−.164	−.216
		(.062)	(.062)	(.070)	(.067)
FEPL (F)	...	...	...	.073	.056
				(.023)	(.022)
Progressive sentiment (P^*)	...	...	.026	...	.022
			(.008)		(.007)
Constant	−6.487	.524	.628	.678	.723
		(.128)	(.118)	(.127)	(.118)
Error variance	.545	.165	.128	.134	.111
R^2†	...	.749	.806	.797	.832

NOTE —Numbers in parentheses are estimates of coefficient standard errors under the assumption of OLS For model II and the combination of models II and IV, the assumption of homoscedasticity is not met and thus estimated standard errors are not unbiased (Heckman 1978)

* For discussion of independent variables, see text

† R^2 is calculated under assumptions of OLS

[23] This procedure is equivalent to the method of solving overidentified models from reduced forms given in Section II.3.

Although these calculations preclude rigorous tests of significance, they suggest that both FEPL and progressive sentiment affected the relative economic standing of blacks in 1959. With neither effect present, the remaining variables explain approximately 75% of the variance in relative wages, whereas FEPL and progressive sentiment each add approximately 5% to explained variance when added separately, and together add 8% explained variance. The results suggest that simply looking at the effect of FEPL alone gives an overestimate of its impact: when progressive sentiment is controlled, the coefficient for FEPL drops by almost 25%, although it remains substantial.

V. CONCLUSION

This article has reviewed alternative formulations of the role of discrete variables in recursive structural equation models. Discrete variables arise in many ways in empirical social science, and structural equation models that are sensitive to these varying substantive contexts are available. In particular, discrete variables may enter structural equations as measures of inherently discrete processes or as indicators of unmeasured continuous variables. When discrete variables are dependent, they are best modeled as having underlying continuous counterparts, but these latent variables may be related either deterministically or stochastically to the observed discrete variables. This article has illustrated the alternative substantive contexts in which the various combinations of discrete variable characteristics may be appropriate. In addition, it has discussed a set of recursive models in which discrete and continuous dependent variables can be considered within single models. Thus, models with discrete variables have the same flexibility as structural equation models with only continuous dependent variables. Finally, this article has shown that methods of path analysis for continuous variables can be applied with little additional difficulty to systems of variables in which some endogenous variables are discrete.

APPENDIX

Methods of Estimation

As shown in Section II, model identification can be achieved typically through solving the structural equations for reduced-form equations, obtaining reduced-form estimates, and using these estimates to derive the structural parameters. For model III this step is unnecessary; structural estimates are estimable directly without resort to the reduced form. Models

I, II, and IV , however, in the absence of other overidentifying restrictions, require not only reduced-form coefficients but also an estimate of the covariances between the disturbances in the two equations of the model. The following discussion reviews procedures for estimating the reduced-form parameters and the disturbance covariance which yield the structural parameters through the steps outlined above. It first reviews methods for estimating single equations with binary response variables and then discusses methods for jointly estimating pairs of equations in which one or both of the dependent variables is discrete. These latter methods provide estimates of covariances among disturbances required for identifying models I, II, and IV.[24]

I. SINGLE-EQUATION ESTIMATION

As is well-known, logit and probit models estimated by maximum likelihood are suitable methods for estimating the effects of independent variables on a dichotomous dependent variable (Cox 1970; Finney 1971; Hanushek and Jackson 1977; Nerlove and Press 1976). These methods are briefly reviewed here. Recall equation (2), for the effects of an independent variable on a latent continuous variable Y^* of which d_y is an observed dichotomous indicator:

$$Y^*_i = \beta + \beta X_i + \epsilon_{y_i} , \qquad (2)$$

where i denotes the ith observation ($i = 1, \ldots , N$) and all other notation is as defined above. Let $c_{y_i} = (\beta_0/\sigma_{\epsilon_y}) + (\beta_1/\sigma_{\epsilon_y})X_i$. Then

$$p(d_{y_i} = 1) = \int_{-\infty}^{c_{y_i}} f\left(\frac{\epsilon_y}{\sigma_{\epsilon_y}}\right) d\epsilon_y ,$$

[24] This discussion emphasizes methods that are computationally feasible at the present time. There is no reliable general computer program for maximum likelihood estimation of structural equations with discrete variables that parallels LISREL for continuous variables (Jöreskog and Sörbom 1978; Jöreskog 1973). The methods for calculating reduced-form equations discussed below include both maximum likelihood and less efficient methods. Maximum likelihood estimation of reduced-form equations, however, does not, in general, yield maximum likelihood estimates of corresponding structural parameters inasmuch as overidentifying restrictions on the structural form of the model are not always imposed and may lead to multiple solutions for the same structural parameter. An alternative strategy, not used here, is to compute tetrachoric correlation coefficients between pairs of discrete variables and biserial correlations between discrete and continuous variables and to estimate structural equation models based on the resulting correlation matrix using LISREL (Jöreskog and Sörbom 1981). This method provides unique estimates of coefficients in overidentified models if the estimated correlation matrix is positive definite. The latter condition is not guaranteed by pairwise estimation of tetrachoric and biserial correlations, nor does this procedure yield correct estimates of coefficient standard errors or of test statistics.

$$= \int_{-\infty}^{c_{y_t}} f(t)dt \ ,$$

and

$$p(d_{y_t} = 0) = \int_{c_{y_t}}^{\infty} f(t)dt \ ,$$

where f is the probability density function for the distribution followed by ϵ_{y_t} and $t = \epsilon_y/\sigma_{\epsilon_y}$. Then if f is the density function of the extreme value distribution,

$$\begin{aligned} p(d_{y_t} = 1) &= \int_{-\infty}^{c_{y_t}} \exp(t)/[1 + \exp(t)]^2 dt \\ &= \exp\left(\frac{\beta_0}{\sigma_{\epsilon_y}} + \frac{\beta_1}{\sigma_{\epsilon_y}} X_t\right) \Bigg/ \left[1 + \exp\left(\frac{\beta_0}{\sigma_{\epsilon_y}} + \frac{\beta_1}{\sigma_{\epsilon_y}} X_t\right)\right] , \end{aligned} \tag{A1}$$

and

$$p(d_{y_t} = 0) = 1 \Bigg/ \left[1 + \exp\left(\frac{\beta_0}{\sigma_{\epsilon_y}} + \frac{\beta_1}{\sigma_{\epsilon_y}} X_t\right)\right] ,$$

which defines the logistic model. Then the likelihood is

$$L = \prod_{t=1}^{N} \left\{ \exp\left(\frac{\beta_0}{\sigma_{\epsilon_y}} + \frac{\beta_1}{\sigma_{\epsilon_y}} X_t\right) d_y \Bigg/ \left[1 + \exp\left(\frac{\beta_0}{\sigma_{\epsilon_y}} + \frac{\beta_1}{\sigma_{\epsilon_y}} X_t\right)\right]\right\} ,$$

and maximum likelihood estimates are obtained by picking values of $\beta_0/\sigma_{\epsilon_y}$ and $\beta_1/\sigma_{\epsilon_y}$ that make L as large as possible (Cox 1970; Hanushek and Jackson 1977).

Alternatively, if ϵ_y follows a normal distribution,

$$\begin{aligned} p(d_{y_t} = 1) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c_{y_t}} \exp\left(\frac{-t^2}{2}\right) dt \ , \\ p(d_{y_t} = 0) &= \frac{1}{\sqrt{2\pi}} \int_{c_{y_t}}^{\infty} \exp\left(\frac{-t^2}{2}\right) dt \ , \end{aligned} \tag{A2}$$

which define the probit model. Then

$$L = \prod_{t}^{N} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c_{y_t}} \exp(-t^2)dt\right]^{d_{y_t}} \left[\frac{1}{\sqrt{2\pi}} \int_{c_{y_t}}^{\infty} \exp(-t^2)dt\right]^{(1-d_{y_t})} ,$$

and maximum likelihood estimates are obtained by picking values of $\beta_0/\sigma_{\epsilon_y}$ and $\beta_1/\sigma_{\epsilon_y}$ that make L as large as possible (Finney 1971; Hanushek and Jackson 1977). Computer programs for estimating logit and probit models by maximum likelihood are widely available (e.g., Baker and Nelder 1978).

II. LOGIT VERSUS PROBIT

For the estimation of single-equation models with dichotomous dependent variables, logit and probit models are virtually interchangeable. In specialized circumstances, mathematical models of scientific phenomena may imply one or the other model (Berkson 1951), but in most applications, logit and probit models differ only in their underlying distributional assumption. This is a trivial difference given the similarity of the logistic and cumulative normal functions (Hanushek and Jackson 1977). Logit models are attractive because they can be derived from log-linear models for frequency data (Fienberg 1980) and provide a closed-form expression for probabilities (compare [A1] and [A2]), but this feature provides no advantage or disadvantage in the analysis of binary data per se.

For multiequation models, in contrast, the choice of logit or probit models is more consequential. As noted above, the main complication in estimating multiequation structural models with discrete endogenous variables is the estimation of the covariance of the disturbances in the reduced-form equations. When the logit model is extended to more than a single dichotomous dependent variable, in its usual form it fails to yield an estimate of the disturbance covariance. In the multinomial logit model, the most common extension of the binary logit model in use, the disturbance covariance is assumed to be zero (McFadden 1974). Naturally, this precludes estimation of the cross-equation dependence implied by the models discussed here.[25] In contrast, the probit model is more easily generalized to a multiequation problem. For a pair of probit equations, the disturbances are bivariate normal, and their covariance can be estimated from the parameter for correlation in the bivariate normal model. Moreover, the assumption of normally distributed disturbances in the probit model facilitates estimation of models with both discrete and continuous dependent variables. Models for the latter typically assume normality of disturbances, and thus two-equation models with one continuous

[25] There are other bivariate distributions than the multinomial that possess logistic marginals and allow disturbances to be correlated. Gumbel (1961) discusses two such distributions, but they have the disadvantage that the correlation is either restricted to a specific value or varies over a limited range. McFadden (1978, 1980) considers a generalization of the extreme value distribution that allows for correlated errors, but this distribution has yet to achieve wide acceptance.

and one discrete response variable can be estimated under the assumption of bivariate normality of the disturbances when the discrete equation is specified as a probit model. In the following discussion, therefore, the multivariate normal model, based on extensions of probit analysis, is employed to estimate the reduced-form error covariance required for the identification of the models discussed above.[26]

The following discussion of the multivariate probit model first presents the likelihood equations for two equations with discrete dependent variables and for two equations when one has a discrete and the other a continuous dependent variable. Then it turns to alternative methods for estimating the coefficients and disturbance covariance for such models when computer software for maximum likelihood estimation for multi-equation systems is unavailable.

III. MAXIMUM LIKELIHOOD FOR TWO-EQUATION MODELS

1. Two Discrete Dependent Variables

Recall that the identification of model I is based on two equations, the structural equation for Y^* and the reduced-form equation for Z:

$$Y^*_i = \beta_0 + \beta_1 X_i + \epsilon_{y_i} , \tag{11}$$

$$Z_i = \gamma_0 + \gamma_1 X_i + \eta_{z_i} , \tag{12}$$

where i denotes the ith observation. The maximum likelihood procedure can be illustrated for this reduced-form model, but the identical estimation problem arises for model II (for the reduced forms for Y' and Z) and for model IV (for reduced forms for Y'_1 and Y'_2). Let Z be a latent continuous variable that is indicated by an observed dichotomous variable d_z. Assume that ϵ_y and γ_z follow a bivariate normal distribution with var $(\epsilon_y) = \sigma_{\epsilon_y}^2$, var $(\eta_z) = \sigma_{\eta_z}^2$, and cov $(\epsilon_y, \eta_z) = \rho\sigma_{\epsilon_y}\sigma_{\eta_z}$. Define $t_y = \epsilon_y/\sigma_{\epsilon_y}$ and $t_z = \sigma_{\eta_z}$ as standardized normal variables. Then the probability density function for t_y and t_z is the bivariate normal density (e.g., Hogg and Craig 1970):

$$g(t_y,t_z) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-(t_y^2 - 2\rho t_y t_z + t_z^2)/(1 - \rho^2)\right] . \tag{A3}$$

[26] This argument implies that when the disturbance covariance is unnecessary to achieve identification, as in the modification of model I in which there is an instrumental variable that affects Y^* but not Z, the logit model may still be used. For an example of this approach, see Duncan and Duncan (1978, pp. 287–96). Nonetheless, when Z is continuous, it remains attractive to estimate both equations under the assumption of normally distributed errors.

For this model there are four possible outcomes on the dependent variables, namely (1) $d_y = d_z = 1$; (2) $d_y = 1$, $d_z = 0$; (3) $d_y = 0$, $d_z = 1$; (4) $d_y = d_z = 0$. Let $c_{y_i} = (\beta_0/\sigma_{\epsilon y}) + (\beta_1 X_i/\sigma_{\epsilon y})$ and $c_{z_i} = (\gamma_0/\sigma_{\eta_z}) + (\gamma_1 X_i/\sigma_{\eta_z})$. Given the bivariate normal model, then

$$p(d_{y_i} = 1,\ d_{z_i} = 1) = \int_{-\infty}^{c_{y_i}} \int_{-\infty}^{c_{z_i}} g(t_{y_i}, t_{z_i})\, dt_z\, dt_y \ ,$$

$$p(d_{y_i} = 1,\ d_{z_i} = 0) = \int_{-\infty}^{c_{y_i}} \int_{c_{z_i}}^{\infty} g(t_{y_i}, t_{z_i})\, dt_z\, dt_y \ ,$$

$$p(d_{y_i} = 0,\ d_{z_i} = 1) = \int_{c_{y_i}}^{\infty} \int_{-\infty}^{c_{z_i}} g(t_{y_i}, t_{z_i})\, dt_z\, dt_y \ ,$$

$$p(d_{y_i} = 0,\ d_{z_i} = 0) = \int_{c_{y_i}}^{\infty} \int_{c_{z_i}}^{\infty} g(t_{y_i}, t_{z_i})\, dt_z\, dt_y \ ;$$

and the likelihood is

$$L = \prod_{i=1}^{N} [p(d_{y_i} = d_{z_i} = 1)]^{d_{y_i} d_{z_i}} \, [p(d_{y_i} = 1,\ d_{z_i} = 0)]^{d_{y_i}(1-d_{z_i})}$$

$$[p(d_{y_i} = 0,\ d_{z_i} = 1)]^{d_{z_i}(1-d_{y_i})} \, [p(d_{y_i} = d_{z_i} = 0)]^{(1-d_{y_i})(1-d_{z_i})} \ .$$

Maximum likelihood estimates are obtained by picking values of $\beta_0/\sigma_{\epsilon y}$, $\beta_1/\sigma_{\epsilon y}$, $\gamma_0/\sigma_{\eta y}$, $\gamma_1/\sigma_{\eta z}$, and ρ that make L as large as possible. The error variances $\sigma_{\epsilon y}^2$ and $\sigma_{\eta y}^2$ can be derived through scale restrictions on Y^* and Z, and thus the disturbance covariance can be obtained from the equation

$$\text{cov}\,(\eta_z, \epsilon_y) = \rho \sigma_{\epsilon y} \sigma_{\epsilon z} \ . \qquad \text{(A4)}$$

2. One Continuous and One Discrete Dependent Variable

Now consider model I again but allow Z to be an observed continuous variable instead of a latent variable. Identical estimation problems arise for model II (for the reduced forms for Y' and Z). As before, define $t_y = \epsilon_y/\sigma_{\epsilon y}$, but define $t_z = (Z - \gamma_0 - \gamma_1 X)/\sigma_{\eta_z}$, which indicates that Z, unlike Y^*, is observed. Again, let $g(t_y, t_z)$ be the bivariate normal density defined above (A3). Then the likelihood is

$$L = \prod_{i=1}^{N} \left[\int_{-\infty}^{c_{y_i}} g(t_y, t_z)\, dt_y \right]^{d_y} \left[\int_{c_{y_i}}^{\infty} g(t_y, t_z)\, dt_y \right]^{(1-d_y)} \ .$$

Maximum likelihood estimates are obtained by picking values of $\beta_0/\sigma_{\epsilon_y}$, $\beta_1/\sigma_{\epsilon_y}$, γ_0, γ_1, σ_{η_z}, and ρ that make L as large as possible. The error variance $\sigma_{\epsilon_y}{}^2$ can be derived through a scale restriction on Y^* (no such restriction is required to identify $\sigma_{\eta_z}{}^2$ since Z is observed). As before, the disturbance covariance is obtained from (A4).

IV. NONLINEAR WEIGHTED LEAST SQUARES ESTIMATORS

The methods just discussed provide efficient estimates of the reduced-form parameters and the disturbance covariances, from which the structural parameters may be derived. These methods are computationally expensive and difficult to implement in the absence of reliable computer software. Most critically, when the number of discrete endogenous variables exceeds two or three, the computation of multivariate normal probabilities becomes intractable inasmuch as each iteration of calculation requires the evaluation for each observation of a probability which is a multiple integral of the same order as the number of discrete endogenous variables (e.g., Muthén 1979). Fortunately, alternative methods are available that hold the promise of allowing consistent (though not efficient) estimation of overidentified structural equation models with discrete variables as well as correct estimation of parameter standard errors and suitable test statistics. These methods, noted briefly here, are discussed more fully elsewhere (Avery, Hansen, and Hotz 1981; Avery and Hotz 1981).

The alternative estimation methods, which employ a nonlinear least squares procedure, use only a subset of the information available about the joint distribution of the discrete endogenous variables that is used in maximum likelihood. In particular, they use the univariate and bivariate moments of the discrete variables but ignore the higher-order moments, thereby avoiding the calculation of higher-order multiple integrals. The estimation strategy is as follows. First, solve the structural equation model for its reduced form and consider the nonlinear functions for each dichotomous dependent variable and for each pair of dependent variables, that is, for the jth dichotomous endogenous variable d_{y_j} ($j = 1, \ldots, p$) (excluding a subscript for individual observations),

$$d_{y_j} = \int_{-\infty}^{\sum_m \Pi_{mj} X_m} \phi(t_j) dt_j + v_j ,$$

and, defining $d_{y_{kl}} = d_{y_k} d_{y_e} (l < k \leq p)$,

$$d_{y_{kl}} = \int_{-\infty}^{\sum_m \Pi_{mk} X_m} \int_{-\infty}^{\sum_m \Pi_{m1} X_m} g(t_k t_l) dt_k dt_l + w_{kl} ,$$

where d_y is a dichotomous variable taking the value 1 or 0, Π_{mj} is the reduced-form parameter for the effect of the mth exogenous variable X_m on the jth endogenous variable d_{y_j}, $\phi(t_j)$ and $g(t_k t_l)$ are the standard univariate and bivariate normal probability density functions, respectively, and v_j and w_{kl} are residuals.

Second, apply a weighted nonlinear least squares criterion to estimate the structural parameters from the system of $p + p(p - 1)/2$ reduced-form equations by minimizing the weighted sums of squares of the residuals v_j and w_{kl} with respect to the structural parameters. The weighting procedure takes account of the differential variances associated with different sums of squared residuals and provides a rigorous method of combining overidentifying restrictions to yield unique estimates. The minimization and the solution of the resulting system of nonlinear equations can be carried out using standard methods for nonlinear estimation (e.g., Goldfeld and Quandt 1972). The procedure leads straightforwardly to estimates of the variances and covariances of the estimated parameters (Avery and Hotz 1981). Although these methods are not yet in wide use, they have been applied in special cases (Avery et al. 1981). Computer software soon to be available will allow their application to the general set of models discussed in this article (Hotz, personal communication, 1982).

REFERENCES

Alwin, Duane F., and Robert M. Hauser. 1975. "The Decomposition of Effects in Path Analysis." *American Sociological Review* 40 (February): 37–47.

Amemiya, Takeshi. 1974. "Bivariate Probit Analysis: Minimum Chi-Square Methods." *Journal of the American Statistical Association* 69 (December): 940–44.

———. 1976. "The Maximum Likelihood, the Minimum Chi-Square and the Nonlinear Weighted Least-Squares Estimator in the General Qualitative Response Model." *Journal of the American Statistical Association* 71 (June). 347–51.

———. 1978. "The Estimation of a Simultaneous Equations Generalized Probit Model." *Econometrica* 46 (September): 1193–1206.

———. 1981. "Qualitative Response Models: A Survey." *Journal of Economic Literature* 19 (December): 1483–1536.

Amemiya, Takeshi, and Frederick Nold. 1975. "A Modified Logit Model." *Review of Economics and Statistics* 57:255–57.

Ashford, J. R., and R. R. Sowden. 1970. "Multi-variate Probit Analysis." *Biometrics* 26 (September): 535–46.

Avery, R. B., L. P. Hansen, and V. J. Hotz. 1981. "Multiperiod Probit Models and Orthogonality Condition Estimation." Discussion Paper 81-12. Chicago: Economics Research Center, National Opinion Research Center, University of Chicago.

Avery, R. B., and V. J. Hotz. 1981. "Estimation of Multiple Indicator Multiple Cause (Mimic) Models with Dichotomous Indicators." Working Paper 60-80-81. Pittsburgh: Graduate School of Industrial Administration, Carnegie-Mellon University.

Baker, R. J., and J. A. Nelder. 1978. *The GLIM System. Release 3. Generalised Linear Interactive Modelling.* Oxford: Royal Statistical Society.

Berkson, Joseph. 1951. "Why I Prefer Logits to Probits." *Biometrics* 7 (December): 327–39

Bielby, William T., and Robert M. Hauser. 1977. "Structural Equation Models." *Annual Review of Sociology* 3:137–61.

Brier, S. 1978. "The Utility of Systems of Simultaneous Logistic Response Equations." Pp. 119–29 in *Sociological Methodology 1979,* edited by Karl F. Schuessler. San Francisco: Jossey-Bass.

Clogg, Clifford. 1980. "New Developments in Latent Structure Analysis." Pp. 215–46 in *Factor Analysis and Measurement in Sociological Research,* edited by D. J. Jackson and E. F. Borgatta. Beverly Hills, Calif.: Sage.

Cox, D. R. 1970. *The Analysis of Binary Data.* London: Methuen.

Davis, James. 1975. "Analyzing Contingency Tables with Linear Flow Graphs: D Systems." Pp. 111–45 in Heise, ed., 1975.

Davis, James, and Susan R. Schooler. 1974. "Non-Parametric Path Analysis—Multivariate Structure of Dichotomous Data When Using the Odds Ratio or Yule's Q." *Social Science Research* 3:267–97.

Duncan, Beverly, and Otis D. Duncan. 1978. *Sex Typing and Social Roles.* New York: Academic Press.

Duncan, Otis D. 1966. "Path Analysis. Sociological Examples" *American Journal of Sociology* 72:1–16.

———. 1975. *Introduction to Structural Equation Models.* New York: Academic Press.

Duncan, Otis D., David L. Featherman, and Beverly Duncan. 1972. *Socioeconomic Background and Achievement.* New York. Seminar

Fienberg, Stephen E. 1975. "Comment" *Journal of the American Statistical Association* 70 (September): 521–23.

———. 1980. *The Analysis of Cross-classified Categorical Data.* 2d ed. Cambridge, Mass.: MIT Press.

Finney, D. J. 1971. *Probit Analysis.* 3d ed. Cambridge. Cambridge University Press.

Goldberger, Arthur S. 1971. "Econometrics and Psychometrics: A Survey of Communalities." *Psychometrika* 36:83–107.

Goldberger, Arthur S., and Otis D. Duncan, eds. 1973. *Structural Equation Models in the Social Sciences.* New York: Seminar.

Goldfeld, Stephen M., and Richard E. Quandt. 1972. *Nonlinear Methods in Econometrics.* Amsterdam: North-Holland.

Goodman, Leo A. 1972. "A General Model for the Analysis of Surveys." *American Journal of Sociology* 77 (May): 1035–86.

———. 1973*a*. "The Analysis of Multidimensional Contingency Tables When Some Variables Are Posterior to Others: A Modified Path Analysis Approach." *Biometrika* 60:179–92.

———. 1973*b*. "Causal Analysis of Data from Panel Studies and Other Kinds of Surveys." *American Journal of Sociology* 78 (March): 1135–91.

———. 1974 "The Analysis of Systems of Qualitative Variables When Some of the Variables Are Unobservable. Part I—a Modified Latent Structure Approach " *American Journal of Sociology* 79 (March): 1179–1259.

———. 1979. "A Brief Guide to the Causal Analysis of Data from Surveys " *American Journal of Sociology* 84 (March): 1078–95.

———. 1981. "Association Models and the Bivariate Normal for Contingency Tables with Ordered Categories." *Biometrika* 68:347–55.

Gortmaker, Steven L. 1979. "Poverty and Infant Mortality in the United States." *American Sociological Review* 44 (April): 280–97

Gumbel, E. J. 1961. "Bivariate Logistic Distributions." *American Statistical Association Journal* 56 (June): 335–49.

Hanushek, Eric A., and John E. Jackson. 1977. *Statistical Methods for Social Scientists.* New York: Academic Press

Hausman, J., and D. Wise. 1978. "A Conditional Probit Model for Qualitative Choice· Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences " *Econometrica* 46 (March). 403–26.

Heckman, James. 1976. "Simultaneous Equation Models with Both Continuous and Discrete Endogenous Variables with and without Structural Shift in the Equations." Pp. 235–72 in *Studies in Nonlinear Estimation,* edited by Stephen M. Goldfeld and Richard E. Quandt. Cambridge, Mass.: Ballinger.

———. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46 (July): 931–59.

Heckman, James, and Burton Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica.* In press.

Heise, David R., ed. 1975 *Sociological Methodology 1976.* San Francisco: Jossey-Bass.

Hogg, Robert V., and Allen T. Craig 1970. *Introduction to Mathematical Statistics.* 3d ed. London· Macmillan.

Jencks, Christopher, et al. 1979. *Who Gets Ahead? The Determinants of Success in America* New York: Basic.

Johnson, Norman, and Samuel Kotz. 1970. *Continuous Univariate Distributions-1.* New York· Wiley.

Jöreskog, Karl 1973. "A General Method for Estimating a Linear Structural Equation System." Pp. 85–112 in Goldberger and Duncan, eds., 1973.

Jöreskog, Karl, and Dag Sörbom. 1978 *LISREL IV User's Guide.* Chicago: National Educational Resources.

———. 1981. *LISREL. Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Version V.* Uppsala· University of Uppsala.

Landes, William. 1968. "The Economics of Fair Employment Laws." *Journal of Political Economy* 76 (July/August): 507–52.

Lazarsfeld, P. F., and N. W. Henry. 1968. *Latent Structure Analysis.* Boston: Houghton Mifflin.

Leik, Robert. 1975 "Causal Models with Nominal and Ordinal Data: Retrospective." Pp. 271–75 in Heise, ed., 1975.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." Pp. 105–42 in *Frontiers in Econometrics,* edited by Paul Zarembka. New York: Academic Press

———. 1976. "Quantal Choice Analysis: A Survey." *Annals of Economic and Social Measurement* 5 (Fall): 363–90.

———. 1978. "Modelling the Choice of Residential Location." *Transportation Research Record* 673:72–77.

———. 1980. "Econometric Models of Probabilistic Choice." Pp. 198–272 in *Structural Analysis of Discrete Data,* edited by Charles F. Manski and Daniel McFadden. Cambridge, Mass.: MIT Press.

McKelvey, Richard D., and William Zavoina. 1975. "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology* 4:103–20.

Manski, Charles F. 1981. "Structural Models for Discrete Data: The Analysis of Discrete Choice." Pp. 58–109 in *Sociological Methodology 1981,* edited by Samuel Leinhardt. San Francisco· Jossey-Bass.

Muthén, Bengt. 1979. "A Structural Probit Model with Latent Variables." *Journal of the American Statistical Association* 74 (December): 807–11.

Nerlove, Marc, and S. James Press 1976. "Multivariate Log-Linear Probability Models for the Analysis of Qualitative Data." Discussion Paper no. 1. Chicago. Center for Statistics and Probability, Northwestern University.

Pearson, K. 1900. "Mathematical Contributions to the Theory of Evolution in the Inheritance of Characters Not Capable of Exact Quantitative Measurement, VIII." *Philosophical Transactions of the Royal Society,* ser. A, 195.79–150.

Rosenthal, Howard. 1980. "The Limitations of Log-linear Analysis." *Contemporary Sociology* 9 (March): 207–12.

Sewell, William H., and Robert M. Hauser. 1975. *Education, Occupation, and Earnings.* New York: Academic Press.

Sewell, William H., and V. P. Shah. 1968. "Social Class, Parental Encouragement, and Educational Aspirations." *American Journal of Sociology* 73:559–72.

Stolzenberg, Ross M. 1979. "The Measurement and Decomposition of Causal Effects in Nonlinear and Nonadditive Models." Pp. 459–88 in *Sociological Methodology 1980,* edited by Karl F. Schuessler. San Francisco. Jossey-Bass.

Yule, G. U. 1900. "On the Association of Attributes in Statistics: With Illustration from the Material of the Childhood Society." *Philosophical Transactions of the Royal Society,* ser. A, 194:257–319.

Zellner, Arnold, and Tong Hun Lee. 1965. "Joint Estimation of Relationships involving Discrete Random Variables." *Econometrica* 33 (April): 382–94.