ENDOGENOUS SWITCHING REGRESSION MODELS FOR THE CAUSES AND EFFECTS OF DISCRETE VARIABLES

Robert D. Mare
and
Christopher Winship

CDE Working Paper 87-32

ENDOGENOUS SWITCHING REGRESSION MODELS FOR THE CAUSES AND EFFECTS OF DISCRETE VARIABLES

Robert D. Mare
University of Wisconsin

Christopher Winship
Northwestern University

January 1987

*This chapter will appear in *COMMON PROBLEMS IN QUANTITATIVE SOCIAL RESEARCH*, edited by J. Scott Long (Sage University Press).

Grants from the National Institute of Aging to Robert D. Mare and Christopher Winship and from the National Institute of Child Health and Human Development (HD 05876) to the Center for Demography and Ecology, University of Wisconsin-Madison supported the preparation of this chapter. The authors are grateful to Meichu D. Chen for expert research assistance and comments on an earlier draft; to Adam Gamoran for advice and the use of his extract of the High School and Beyond data; and to J. Scott Long and Judith A. Seltzer for helpful comments.

INTRODUCTION

A common research problem is to assess the consequences for individuals of their social roles, statuses, or group memberships. For example, labor market analyses assess the effects on individuals' earnings of their occupations, labor market sectors, social classes, or educational credentials. Education research examines the effects of students' placement in tracks on their academic achievement. Evaluation studies estimate the effects of participation in social programs on social policy outcomes. These analyses attempt either to establish the effects of positions or statuses on outcomes or to see whether other variables differ in their effects on the outcomes across positions. This chapter describes some models for these kinds of analyses that improve upon more commonly use models.

A typical approach to data on social positions and their outcomes is the analysis of covariance (ANACOVA) (e.g., Long and Miethe, this volume). This approach assumes that, once other measured variables that affect the outcome (dependent) variable are taken into account, the process by which individuals are sorted into positions is independent of factors influencing the outcome variable itself. That is, no unmeasured variables affect both the dependent variable and the variables that indicate an individual's position. Such strong assumptions are often untenable. Individuals may choose (or be assigned to) positions based on the expected consequences of those positions for the outcome of interest. For example, school administrators or parents may assign children to school tracks according to the academic benefit that the children will derive from one track versus another. Individuals may decide whether or not to attend college on the basis of the perceived benefit of college relative to that of not attending college. Individuals may selfselect themselves into social programs in part because of the benefits of program participation.

In these examples, positions occupied affect outcomes and the outcomes associated with positions affect the assignment of persons to positions. Thus outcomes and positions are simultaneously related. The simultaneity creates a selection bias in that the assignment of persons to positions is not random, even controlling for measured independent variables, but instead is a function of the outcome variable. Omitted variable and simultaneity problems imply that ANACOVA models estimated by OLS yield inconsistent estimates of the effects of social positions and of other independent variables in the models.

This chapter describes models for the joint determination of a discrete outcome, which denotes individuals' statuses or positions, and another outcome--either discrete or continuous--that may be affected by the status or position. These models, sometimes called "endogenous switching models," relax the assumptions of ANACOVA by allowing for the *joint* determination of the discrete variables and the outcomes that they affect. The models enable one to (1) model both the allocation of persons to positions and the effects of positions on other outcomes; (2) estimate the degree to which common, unmeasured variables affect both the outcome and the classification variables; (3) obtain estimates of the effects of other variables *within* levels of the classification variables that take account of potential selection biases; and (4) estimate the impact of the classification regime (e.g., tracking, market segmentation, program treatments) by simulating how individuals would fare had they entered different positions from those that they in fact occupy.

Much of the material discussed in this chapter is covered in more technical form by Maddala (1983:257-90), Amemiya (1986:360-411), and the literature reviewed by them. The endogenous switching model is closely related to two classes of models, namely sample selection (e.g., Berk, 1983) and dummy endogenous variable models (Heckman, 1978). Although the goal of this chapter is to provide a didactic account of endogenous switching models, some of the structural forms of the model are presented here for the first time.

This chapter is restricted to a discussion of problems where the assignment of observations to categories, as well as the realized outcomes for those categories are both *observed* endogenous variables. Another class of models, not discussed here, assumes an *exogenous* but *un*observed classification. These latter models, also called switching regressions, rely on within-sample heterogenity in the functional forms that relate observed exogenous and outcome variables [see, e.g., Goldfeld and Quandt (1973), Johnston (1984:407-409), and Maddala (1977:394-396)].

ENDOGENOUS SWITCHING MODELS FOR POSITIONS AND THEIR OUTCOMES

We begin with a general model of the sorting of persons to positions and of the effects of position on outcomes. Then we show how the general model can be restricted to incorporate assumptions about the rules that govern the assignment of persons to positions. We consider three specific cases. One case, the "Ascription Model," assumes that individuals are assigned to positions solely on the basis of their observed characteristics and not on the basis of their expected outcomes. A second, the "Maximization Model," assumes that persons are optimally assigned, that is, they enter the positions where their expected outcomes are most favorable. The final case, the "Quota Model," assumes that one of the positions is "dominant" in the sense that persons are assigned according to how well they would be expected to do if they entered that position, and their expected outcomes in other positions are irrelevant to the assignment decision.

"Positions" and "Outcomes"

In the following discussion "positions" denote a classification of two or more roles, statuses, or other categories into which persons are sorted. "Outcomes" denote the rewards, achievements, or consequences associated with alternative positions. We treat these outcomes as continuous variables, but similar models are available for discrete outcomes. In most applications the outcomes may be regarded as both the *ex post* rewards that depend on entry into positions, and also *ex ante* incentives for persons to choose one position over others. Position and outcome, therefore, may be *jointly* determined.

Decision-Making and Rationality

In many applications of these models, the allocation of persons to positions can be viewed as the decisions of individual actors. Some of our discussion applies this view. The *models*, however, do not depend on the assumption of rational decision-making that one might find in applications of the models that derive explicitly from economic or other behavioral theories. Rather, the models are consistent with the view that the sorting of persons to positions is the outcome of the simultaneous actions of *many* persons, such as employers, workers, relatives, etc. Alternatively, the models may be viewed as altogether independent of the concepts of decision-making and choice. Some decision-making processes can be formalized and empirically tested against more flexible versions of the model. Thus, assumptions about decision-making are testable hypotheses rather than assumptions inherent in the general model.

The General Model

Denote persons by i ($i = 1,\ldots,I$) and positions by j ($j = 1,2$). Let d_i

be a dichotomous variable that equals 1 for persons entering position 1 and 0 for persons entering position 2. For the *ith* person let Y_i denote the outcome for the *ith* person and X_{ki} denote the value on the *kth* measured independent variable ($k = 1,\ldots,K$) that may affect position assignment or outcome within positions. Assume that the X_k are predetermined with respect to both the allocation of persons to positions and outcome within the position. Assume further that $X_{1i} = 1$ for all individuals so that it enters the model as a constant. Under these assumptions, the usual ANACOVA model is

$$(1) \quad Y_i = \alpha_0 d_i + \sum_k \alpha_k X_{ki} + \epsilon_i$$

where ϵ_i denotes a stochastic disturbance and the α_k ($k = 0,\ldots,K$) denote parameters to be estimated. Under suitable assumptions about ϵ_i, this model can be estimated by OLS. It can, moreover, be generalized to allow the effects of the X_k to depend on d. This generalization can be written with interaction terms added to (1) or as two separate equations, say,

$$(2) \quad Y_{1i} = \sum_k \beta_{1k} X_{ki} + \epsilon_{1i},$$

$$(3) \quad Y_{2i} = \sum_k \beta_{2k} X_{ki} + \epsilon_{2i},$$

where Y_1 and Y_2 denote outcomes in positions 1 and 2 respectively, ϵ_1 and ϵ_2 denote stochastic disturbances, and the β_{1k} and β_{2k} denote parameters to be estimated. Under this formulation the outcome becomes two variables for each individual. Of course, one typically observes only the outcome associated with the position that each individual in fact holds. As discussed below, however, we consider the possibility that both observed outomes and the hypothetical outcomes that persons would experience had they entered different positions affect their assignments to positions.

In restricted forms of the model some of the X_k may be excluded from one or both of equations (2) and (3). In addition, some variables may have the same effect on the outcome in position 1 as in position 2, that is $\beta_{1k} = \beta_{2k}$ for some k. If the effects of the independent variables are identical in the two positions for all independent variables except the constant, that is, $\beta_{1k} = \beta_{2k}$ for k>1, then the net effect of position on outcome is $\beta_{11} - \beta_{21}$. If, on the other hand, the effects of the X_k vary across positions, there is no single position effect. Rather the position effect is conditional upon the values of the X_k. Thus, the model allows for the additive or interactive effects of position on outcome that are represented in common ANACOVA models.

Now consider an extension of the model, which represents the process by which persons are allocated to positions. Let persons have latent scores Z_i which index their likelihood of assignment to position 1. The probability of assignment to position 1 is $P(Z_i > 0)$. Let the relative chances that a person is assigned to position 1 or 2 be a function of both the same predetermined factors X_k that affect outcomes in positions, and also the (expected) outcomes of the positions themselves. This extension of the model can be written:

(4) $$Z_i = \sum_k \gamma_k X_{ki} + \eta_1 Y_{1i} + \eta_2 Y_{2i} + \zeta_i ,$$

where η_1, η_2, and the γ_k denote parameters and ζ denotes a stochastic disturbance. Throughout, we assume that ζ is uncorrelated with ϵ_1 and ϵ_2 in equations (2) and (3).

For estimating and interpreting the general model and its restricted forms, it is useful to put (4) into its reduced form, that is:

(5) $$Z_i = \sum_k \pi_k X_{ki} + \epsilon_{3i} ,$$

where, from (2) and (3), $\pi_k = \eta_1 \beta_{1k} + \eta_2 \beta_{2k} + \gamma_k$ and $\epsilon_{3i} = \eta_1 \epsilon_{1i} + \eta_2 \epsilon_{2i} + \zeta_i$. Thus (2), (3), and (4) describe the structural form of a model for the allocation of persons to positions and the effects of positions on outcomes; and (2), (3), and (5) describe the corresponding reduced form of the model.

To complete the model, it remains to specify the structure of the reduced form disturbances. Assume that ϵ_1, ϵ_2, and ϵ_3 follow a trivariate normal distribution with $Var(\epsilon_1) = \sigma_1^2$, $Var(\epsilon_2) = \sigma_2^2$, $Var(\epsilon_3) = \sigma_3^2$, $Cov(\epsilon_1,\epsilon_2) = \sigma_{12}$, $Cov(\epsilon_1,\epsilon_3) = \sigma_{13}$, and $Cov(\epsilon_2,\epsilon_3) = \sigma_{23}$. The disturbance in the structural equation for allocating persons to positions, ζ, is uncorrelated with the disturbances in the outcome equations, ϵ_1 and ϵ_2. The reduced form disturbance ϵ_3, however, is generally correlated with ϵ_1 and ϵ_2. Equations (1) and (2) define regression models, albeit models that must be estimated in a way that takes account of the correlated disturbances. Because ϵ_3 is normally distributed, (5) is a probit model (e.g., Aldrich and Nelson, 1984). The normality assumption makes estimation of the full model feasible and enables us to make useful calculations on the basis of estimates from the model.

Note again that the distributions of not only Z but also Y_1 and Y_2 are defined for the *entire population*, not just for persons who in fact enter position 1 or position 2. For example, if $\bar{X}_k$ is the population mean of X_k, then $\Sigma\beta_{1k}\bar{X}_k$ is the expected level of outcome for a *random* sample of the entire population when placed in position 1. Of the parameters that we have mentioned, however, all except σ_{12} and σ_3 can be estimated from data on the X_{ki}, on d_i, and on one measure on either Y_1 *or* Y_2 for each person. To estimate σ_{12} and σ_3 requires additional information, and, in practice, we assume values for these parameters (see discussion of identification and estimation below).

As written, all of the X_k affect each of the three endogenous variables, but some of the X_k may be excluded from some of the equations. For example, some variables may affect outcome but not position or vice versa.

Alternatively, some variables may affect the outcome in one position but not in the other. When it is substantively plausible to exclude some of the X_k from some of the equations, the resulting reduction in possible multicollinearity may make it is easier to obtain stable estimates of parameters than when all X_k appear in all three equations.

Unlike the ANACOVA model, however, the endogenous switching regression model includes the effects of nonrandom selection into positions. The ANACOVA model assumes that unmeasured determinants of position placement, ϵ_3, are uncorrelated with unmeasured determinants of outcome within positions, ϵ_1 and ϵ_2. If this assumption holds, then ordinary least squares (OLS) estimates of position effects and of within position effects of the other X_k on outcome are consistent. If, however, the assumption is false then OLS estimates of the β_{jk} in (1) and (2) are *in*consistent because of the nonrandom selection of persons into positions (even conditional upon the observed X_k). In the model discussed here, the covariances between the disturbances in the equation predicting allocation of persons to positions, ϵ_3, and in the outcome equations, ϵ_1 and ϵ_2, adjust for this potential inconsistency. The covariances σ_{13} and σ_{23}, moreover, show the degree and direction of nonrandom selection of persons to positions. This model, therefore, represents the "structural" or population-level effects of position on outcome, rather than just a description of differences in outcomes between observed position 1 and position 2 samples.

Selection of Persons into Positions

Using the endogenous switching model, we can investigate the direction and degree of nonrandom selection of persons to positions and the selection biases that are implicit in OLS estimates of position effects. We can also simulate how persons would fare if placed in alternative positions. These analyses use formulas for the means of censored normal distributions (e.g., Maddala, 1983:365-7).

Assume that we have the model defined by equations (2)-(5). Let $\hat{Z}_i$ denote the predicted value of Z from the X_k for the *ith* individual, that is, $\hat{Z}_i = \Sigma\pi_k X_{ki}$. The expected outcome for the *ith* person if they enter position 1, that is, persons for whom $Z_i > 0$, is

$$(6) \quad E(Y_{1i} \mid Z_i > 0) = \sum_k \beta_{1k} X_{ki} + E(\epsilon_{1i} \mid Z_i > 0;\ X_{1i}, \ldots, X_{iK})$$
$$= \sum_k \beta_{1k} X_{ki} + (\sigma_{13}/\sigma_3)\{[\phi(\hat{Z}_i)]/[\Phi(\hat{Z}_i)]\},$$

where ϕ denotes the normal probability density function and Φ denotes the cumulative normal probability function. This expression implies that the expected outcome of persons who are placed in position 1 is a function of both their measured independent variables, $\Sigma\beta_{1k}X_{ki}$, and also the expected value of their unmeasured characteristics. The latter quantity is a function of the covariance between outcome in position 1 and the chances of assignment to position 1 (σ_{13}) and (a function of) the probability of assignment to position 1 $\{[\phi(\hat{Z}_i)]/[\Phi(\hat{Z}_i)]\}$. Thus, the outcome of persons *observed* in position 1, $E(Y_{1i} \mid Z_i > 0)$, is a biased estimate of their expected outcome for position 1 where the bias is equal to the second term on the right side of (6). To estimate the expected outcome for an individual in position 1, we need to know both $\Sigma\beta_{1k}X_k$ and also information about the probability of selection into position 1 as well. This is an application of well-known ideas about "sample selection bias" (e.g., Berk, 1983).

The expected outcome for a person observed in position 2 is

$$(7) \quad E(Y_{2i} \mid Z_i < 0) = \sum_k \beta_{2k} X_{ki} + E(\epsilon_{2i} \mid Z_i < 0;\ X_{1i}, \ldots, X_{Ki})$$
$$= \sum_k \beta_{2k} X_{ki} - (\sigma_{23}/\sigma_3)\{[\phi(\hat{Z}_i)]/[1-\Phi(\hat{Z}_i)]\}.$$

The disturbance covariances, σ_{13} and σ_{23}, measure the direction and degree of nonrandom selection into positions. In particular, equation (6) shows that if $\sigma_{13} > 0$, then there is positive selection of persons to position 1, that is, net of the effects of the X_k, Y_1 is higher for persons who enter position 1 than it would be for the average person in the population. Conversely, $\sigma_{13} < 0$ indicates negative selection into position 1. Equation (7) shows that $\sigma_{23} < 0$ indicates positive selection of persons to position 2, whereas $\sigma_{23} > 0$ implies negative selection.

As pointed out by Willis and Rosen (1979) and Maddala (1983:258), two combinations of signs of σ_{13} and σ_{23} are of particular interest. First, suppose $\sigma_{13} > 0$ and $\sigma_{23} < 0$. This is the case of positive selection in both positions. That is, net of the effects of the X_k, persons who enter position 1 do better than the population average for position 1 and persons who enter position 2 do better than the population average for position 2. Conversely, were persons who enter position 1 placed in position 2 instead, they would do *worse* than the persons who in fact enter position 2. This accords with a notion of two dimensions along which persons can be assessed, each of which is best suited to one of the positions.

Second, suppose $\sigma_{13} > 0$ and $\sigma_{23} > 0$. In this case, persons who enter position 1 do better than the population average for that position, but persons who enter position 2 do worse than the population average for that position. Conversely, were persons who enter position 1 to be placed in position 2 instead, they would do better than the persons who in fact entered position 2. This case implies a single dimension on which persons who enter position 1 are more advantaged than persons who enter position 2.

Note that σ_{13} and σ_{23} are parameters to be estimated rather than assumed quantities. Thus, under the assumptions of the model, we can assess which of the two cases is closer to the truth. (A third case, where $\sigma_{13} < 0$ and $\sigma_{23} < 0$, mirrors the second case, that is, a single latent dimension on which persons who enter position 2 are more advantaged than persons who enter position 1. Finally, there is the logically possible, but perverse case where $\sigma_{13} < 0$ and $\sigma_{23} > 0$, which implies negative selection to both positions.)

Alternative Position Assignments

In addition to illustrating the relationship between population and observed outcome levels for the two positions, the endogenous switching regression model provides a way of calculating the expected level of outcome if persons are assigned to positions other than the ones they in fact entered. The expected outcome of position 1 persons had they been placed in position 2 instead is

$$(8)\quad E(Y_{2i} \mid Z_i > 0) = \sum_k \beta_{2k} X_{ki} + (\sigma_{23}/\sigma_3)\{[\phi(\hat{Z}_i)]/[\Phi(\hat{Z}_i)]\}.$$

The expected outcome of position 2 persons had they been placed in position 1 is

$$(9)\quad E(Y_{1i} \mid Z_i < 0) = \sum_k \beta_{1k} X_{ki} - (\sigma_{13}/\sigma_3)\{[\phi(\hat{Z}_i)]/[1-\Phi(\hat{Z}_i)]\}.$$

These ideas and formulas can be illustrated with numerical examples. Consider two persons, one observed in position 1, the other in position 2 who have identical values on X_k. Let $\sigma_3 = 1$. First, suppose $\Sigma\beta_{1k}X_k = 2$, $\Sigma\beta_{2k}X_k = 1$, $Z = 1$, $\sigma_{13} = 0.5$, and $\sigma_{23} = -0.5$. Given the signs of the covariances, we expect that both persons, if placed in the position in which they are *not* observed, would fare worse than the persons in fact do. Applying equations (6)-(9), we see that that is indeed the case:

$$E(Y_1 \mid Z > 0) = E(\text{Pos } 1 \mid \text{Observed in Pos } 1) = 2.14,$$

$E(Y_2 \mid Z < 0) = E(\text{Pos } 2 \mid \text{Observed in Pos } 2) = 1.76$,

$E(Y_2 \mid Z > 0) = E(\text{Pos } 2 \mid \text{Observed in Pos } 1) = 0.86$,

$E(Y_1 \mid Z < 0) = E(\text{Pos } 1 \mid \text{Observed in Pos } 2) = 1.24$.

In a second example, $\Sigma\beta_{1k}X_k = 2$, $\Sigma\beta_{2k}X_k = 1$, $Z = 1.41$, $\sigma_{13} = 1.06$, and $\sigma_{23} = 0.35$. This illustrates the case where $\sigma_{13} > 0$ and $\sigma_{23} > 0$. Thus the person observed in position 1 would achieve a higher outcome than the one observed in position 2 irrespective of which position he or she were placed in. From equations (6)-(9), we compute the following:

$E(Y_1 \mid Z > 0) = E(\text{Pos } 1 \mid \text{Observed in Pos } 1) = 2.17$,

$E(Y_2 \mid Z < 0) = E(\text{Pos } 2 \mid \text{Observed in Pos } 2) = 0.35$,

$E(Y_2 \mid Z > 0) = E(\text{Pos } 2 \mid \text{Observed in Pos } 1) = 1.06$,

$E(Y_1 \mid Z < 0) = E(\text{Pos } 1 \mid \text{Observed in Pos } 2) = 0.03$.

These examples show some alternative assessments of the effects of the sorting of persons to positions. In both examples, persons fare better in the positions where they are observed than in the alternative position, implying that sorting is "optimal" in the sense of placing persons where they will do the best. This result, however, is empirical and is *not* required by the model.

The switching model also allows for a benign or critical judgment of the impact of sorting on group differences. In the first example, the observed difference in outcome between the two positions (2.14 - 1.76 = 0.38) is less than the population difference (2 - 1 = 1), but in the second example, the observed difference (2.17 - 0.35 = 1.82) is larger. Relative to random allocation of persons to positions, therefore, the first example suggests that the observed system of allocation reduces group differences whereas the second shows that it enhances them. These results show the potential richness of the model for understanding the effects of the process of sorting persons to positions on levels and differentials in outcome.

These models can show the potential impact on *marginal* individuals were they to move from one position to another. In the aggregate, they suggest how average levels of the outcome variable would change in their levels and distribution for groups under alternative assignments to the ones observed in the sample. One must be cautious, however, in making such extrapolations from the models to hypothetical systems of assignments. In large populations, movements by small numbers of individuals are unlikely to affect the structure of assignment and outcome for the population as a whole. If, however, substantial portions of the population were reassigned, the overall system and the appropriate model could change, thereby invalidating inference from the orginal model. These models, therefore, can suggest the impact of a system of sorting persons to positions and outcomes, but they are not a substitute for historical or comparative data on alternative systems.

Structural Models: I. Ascription Model

We can now further specify the process by which persons are allocated to positions, and the implications of that process. In the following discussion we consider three models, the Ascription Model, the Maximization Model and the Quota Model. In all three models, one or several decision makers assign persons to positions. The models differ in the principle that governs decision making. Each hypothesizes a *structural form* of the switching regression model that implies restrictions on the general model.

Consider first the case where persons are assigned to positions solely on the basis of their observed (predetermined) characteristics. That is, their expected outcomes Y_1 and Y_2 do not affect their assignment to positions, once their measured characteristics X_k are taken into account. This model, which we term the Ascription Model, amounts to placing the restriction $\eta_1 = \eta_2 = 0$ in (4). Thus $\epsilon_3 = \zeta$, which is uncorrelated with ϵ_1 and ϵ_2. To test the validity of this model it suffices to estimate the reduced form of the general model [equations (2), (3), and (5)] and a restricted model in which $\sigma_{13} = \sigma_{23} = 0$. If these models are estimated by maximum likelihood (see below), then, since the latter model is nested within the former, the restriction can be tested using a likelihood ratio test (using a χ^2 statistic with two degrees of freedom). If the Ascription Model holds, then equations (2) and (3) can be estimated by OLS.

Structural Models: II. Maximization Model

In the Maximization Model persons are allocated to the positions where their expected outcome is highest. That is, the *ith* person enters position 1 if $Y_{1i} > Y_{2i}$ and position 2 if $Y_{2i} > Y_{1i}$. This model assumes that persons could not fare any better were they assigned to an alternative position to the one they in fact entered. It does not dictate, however, whether *differences* between groups of persons would be larger or smaller under alternative position assignments than those that occur.

This model is also a special case of the general model and can be written as a set of mathematical restrictions. In particular, in equation (4), let $\eta_1 = -\eta_2 = \eta$ and $\gamma_k = 0$ for $k > 1$, yielding

$$(10) \qquad \begin{aligned} Z_i &= \gamma_1 + \eta(Y_{1i} - Y_{2i}) + \zeta_i \\ &= \gamma_1 + \eta[(\sum_k \beta_{1k} X_{ki} + \epsilon_{1i}) - (\sum_k \beta_{2k} X_{ki} + \epsilon_{2i})] + \zeta_i \\ &= \sum_k \pi_k X_{ki} + \epsilon_{3i} \end{aligned}$$

where $\pi_k = \eta(\beta_{1k} - \beta_{2k})$ for $k > 1$, $\pi_1 = \gamma_1 + \eta(\beta_{11} - \beta_{21})$, and $\epsilon_{3i} = \eta(\epsilon_{1i} - \epsilon_{2i}) + \zeta_i$. The parameter η in this model is the effect of the difference in expected outcomes between positions 1 and 2 on the chances that a person will be assigned to position 1. The π_k parameters in equation (10), unlike the corresponding parameters in equation (5), are proportional to the differences between the corresponding β_k parameters for outcome in positions 1 and 2. The model also implies restrictions on the covariance matrix of disturbances in equations (2), (3), and (5). That is, $\sigma_{13} = \eta(\sigma_1^2 - \sigma_{12})$, and $\sigma_{23} = \eta(\sigma_{12} - \sigma_2^2)$. These restrictions imply that it is possible to estimate σ_{12}, a parameter that is unidentified in the general model. The restrictions also require that $\sigma_{13} - \sigma_{23} > 0$, which, provided $\eta > 0$, eliminates the perverse outcome that persons are negatively selected into both positions. Conversely, it agrees with the key assumption of the Maximization Model, that persons enter the position where they get the best reward.

If the constraints of the Maximization Model hold, then one can make similar interpretations of the σ_{13} and σ_{23} parameters and calculations of expected outcomes under alternative position assignments to those discussed above.

Structural Models: III. Quota Model

In the Quota Model persons are allocated to positions according to the availability of vacancies in position 1 and to their expected qualifications for the position. Suppose that decision makers seek to fill position 1 with the *top* of the population as defined by its expected outcome in position 1. Conversely, they assign to position 2 persons who are expected to fall in the bottom of the outcome distribution if they enter position 1. Unlike in the Maximization Model, persons' expected outcomes in position 2 are irrelevant to the allocation of persons to positions.

The Quota Model can be specified as a set of restrictions on equation

(4), namely $\eta_2 = \gamma_k = 0$ for $k > 1$, that is

$$Z_i = \gamma_1 + \eta_1 Y_{1i} + \zeta_i,$$

which implies a reduced form of

$$Z_i = \gamma_1 + \sum_k \eta_1 \beta_{1k} X_{ki} + \eta_1 \epsilon_{1i} + \zeta_i$$

$$= \sum_k \pi_k X_{ki} + \epsilon_3,$$

where $\pi_1 = \eta_1 \beta_{11} + \gamma_1$, $\pi_k = \eta_1 \beta_{1k}$ for $k > 1$, and $\epsilon_{3i} = \eta_1 \epsilon_{1i} + \zeta_i$. This model implies that all coefficients in equations (2) and (5) except for the constant term should be proportional to each other and that $\sigma_{13} = \eta_1 \sigma_1^2$. As for the other two restricted models, the validity of the restrictions implied by the Quota Model can be tested by computing the loss of fit of the general model when the restrictions are imposed. But the Quota Model is not nested within the Maximization Model because the latter imposes restrictions upon the parameters of the position 2 outcome equation that are not imposed in the former.

Partial Foresight

Our discussion of the Maximization and Quota Models has assumed that the assignment of persons to positions is based on complete knowledge of Y_1 and Y_2, an assumption that may often be unrealistic. For example, in deciding whether or not to continue in school, individuals may know only roughly the benefit that they will derive from staying in school. In evaluating the consequences of possible actions, therefore, they rely on predicted, rather than actual values of the outcomes.

One way to model this is to assume that persons are assigned to positions on the basis of values of Y_1 and Y_2 predicted from the X_k's for those persons. That is, the person(s) carrying out the assignment know as much as the social researcher, but no more. Then the structural equation for position assignment under the general model (4) becomes:

$$(11)\qquad Z_i = \gamma_1 + \eta_1\hat{Y}_{1i} + \eta_2\hat{Y}_{2i} + \zeta_i$$

where $\hat{Y}_{1i} = \Sigma\beta_{1k}X_{ki}$ and $\hat{Y}_{2i} = \Sigma\beta_{2k}X_{ki}$. The reduced form is

$$(12)\qquad Z_i = \sum_k(\eta_1\beta_{1k} + \eta_2\beta_{2k})X_{ki} + \zeta_i$$

$$= \sum_k \pi_k X_{ki} + \epsilon_{3i}$$

In this specification, ϵ_3 is uncorrelated with ϵ_1 and ϵ_2 in equations (2) and (3). If statistical tests indicate that $\sigma_{13} = \sigma_{23} = 0$, then the outcome equations can be estimated by OLS. When the restrictions on the π_k implied by the Maximization or Quota Models hold, however, it is better to estimate equations (2), (3) and (11) simultaneously (see below).

More Than Two Positions

On can extend the models to include more positions. For three positions, for example, a third outcome equation, parallel to equations (2) and (3) can be added. The sorting equation (5) becomes a trinomial probit equation instead of a binary probit. Then two latent variables, say Z_1 and Z_2, index individuals' relative suitability for the three positions. A special case of the multi-position model occurs when the positions are ordered. For example, one can model decisions to continue in school and the rewards associated with alternative amounts of schooling as a choice equation and a set of outcome equations associated with each level of schooling (Willis and Rosen, 1979). For two schooling levels, equations (2)-(5) are such a model. For three or more levels, additional outcome equations are required for additional schooling levels, as in the case of the general polytomous model (Garen, 1984). The sorting equation (5), however, is then a single *ordered* probit equation (e.g., Winship and Mare, 1984), which predicts schooling as an ordered variable that indexes a single latent variable Z. Although standard computer packages do not readily estimate this model, it is easier to estimate than the general polytomous model.

ESTIMATION

This section first discusses identification issues raised by the endogenous switching model. Then it reviews methods of estimation, including multi-stage consistent estimation and maximum likelihood estimation. Finally, it discusses special estimation issues that are raised by restricted forms of the general model.

Identification

Under the assumption of trivariate normality, the general model defined by equations (2)-(5) and its associated covariance matrix requires two restrictions to be identified. Because only one of Y_1 or Y_2 is observed, no sample information is available to identify the covariance between the disturbances in the two outcome equations, σ_{12}. Because the latent variable Z is only indicated by the binary outcome of whether individuals are in position 1 or position 2, its scale, which is determined by the disturbance variance σ_3 is also unidentified. With regard to σ_{12}, it is necessary to assume a value for this parameter or to impose additional restrictions such as those of the Maximization Model (see above). In the general model, however, estimates of the other parameters in the model do not depend on σ_{12}.

Identifying σ_3 is the usual problem of scale identification in binary regression models (e.g., Winship and Mare, 1983:73-5; 1984:517). To identify σ_3 in the general model one must adopt a normalization rule, such as $\sigma_3 = 1.0$ or Var(Z) = 1.0. These rules affect the estimated parameters for equation (3) and σ_{13} and σ_{23} by a factor of proportionality but do not affect the remaining parameters or the fit of the model.

Apart from the restrictions just discussed, no additional assumptions are needed to identify the β and π parameters. In practice, however, the precision with which the β's are identified may considerably improve if some of the X_k do not appear in some of the equations. When no such exclusion restrictions are available, the identification of σ_{13} and σ_{23} depends on the nonlinearity implicit in the normality assumption. Under this circumstance, the correlations and standard errors of the estimates of σ_{13}, σ_{23}, β_{1k}, β_{2k}, and π_k may be large. Of course, the exclusion restrictions should derive from substantive reasoning rather than *ex post* inspection of empirical results.

Multi-Stage Consistent Estimation

Multi-stage estimators are valuable because they can be implemented with standard programs that perform regression and probit analysis, and they, unlike many maximum likelihood estimators, intuitively demonstrate the workings of the model. In this section we describe one multi-stage estimation routine for the general endogenous switching regression model. Depending on the special features of the model, other routines may also be feasible. Maddala (1983: Ch. 8) discusses multi-stage estimators for related models.

Equations (2), (3), and (5) are two linear regression equations and a dichotomous response equation predicting the probability of entering position 1. Assume that we have a random sample of persons, some of whom are in position 1, the remainder of whom are in position 2. Y_1 is observed only for persons in position 1 and Y_2 is observed only for persons in position 2. Which position an individual has entered is known for the entire sample. The model can be estimated as follows:

(A) Assuming that σ_3=1, estimate the π_k parameters in equation (5) by probit analysis.

(B) Use the estimated parameters to calculate predicted values $\hat{Z}_i$ and the quantities $\phi(\hat{Z}_i)$ and $\Phi(\hat{Z}_i)$ for each individual.

(C) With the estimates from step (B) in hand, compute the following two ratios:

$$\lambda_{1i} = \phi(\hat{Z}_i)/\Phi(\hat{Z}_i); \quad \lambda_{2i} = -\phi(\hat{Z}_i)/[1-\Phi(\hat{Z}_i)].$$

(D) Modify equations (2) and (3) by including λ_1 and λ_2 respectively as regressors, that is

$$(2^*) \quad Y_{1i} = \sum_k \beta_{1k} X_{ki} + \sigma_{13}\lambda_{1i} + \nu_{1i},$$

and

$$(3^*) \quad Y_{2i} = \sum_k \beta_{2k} X_{ki} + \sigma_{23}\lambda_{2i} + \nu_{2i},$$

where ν_{1i} and ν_{2i} are stochastic disturbances that are uncorrelated with the X_k, λ_1, and λ_2. Estimate (2^*) and (3^*) by OLS over the subsamples for which Y_1 or Y_2 are observed.

Step (D) yields consistent estimates of not only β_{1k} and β_{2k}, but also σ_{13} and σ_{23}, which are the coefficients on λ_1 and λ_2 respectively. This procedure is tantamount to the commonly-applied two-stage procedure for correcting regressions for sample selection bias (e.g., Berk, 1983), except that it adjusts for nonrandom selection into two samples instead of one. It shows that the potential inconsistency in OLS estimation of (2) and (3) is a function of the degree of correlation between (functions of) the probability that an individual is selected into position 1 or position 2 (λ_1 or λ_2) and the X_k.

Since λ_1 and λ_2 are functions of $\hat{Z}$, which is a function of the X_k, the above discussion implies that collinearity between λ_1, λ_2, and X_k may be a serious problem in obtaining precise estimates of the parameters. Thus when substantive reasoning implies that some of the X_k can be excluded from some of the equations, collinearity may be substantially reduced.

Steps (A)-(D) consistently estimate the parameters but not the standard errors of the β_{1k} and β_{2k}, a result of heteroskedasticity in (2^*) and (3^*). To obtain consistent estimates of the standard errors, these equations must be estimated by weighted least squares instead of OLS. Maddala (1983:225-227, 252-256) provides guidelines for constructing the weights.

Maximum Likelihood Estimation

Equations (2), (3), and (5), their disturbance covariance matrix, and restricted forms of this model can also be estimated by maximum likelihood. Maximum likelihood is usually the preferred method for these models because it is feasible with available software, provides efficient estimates, allows restrictions to be applied, and permits the construction of likelihood ratio tests of the restrictions.

Maximum likelihood estimation consists of specifying the probability of obtaining the observed data in terms of the parameters to be estimated and picking parameter values that make the probability as large as possible. Again assume a random sample of individuals, some of whom are in position 1 and the rest of whom are in position 2. The likelihood is:

$$(13)\quad L(Y_1,Y_2,d_i) = \prod_i \{(d_i)[f(Y_1 \mid d_i=1)] + (1-d_i)[f(Y_2 \mid d_i=0)]\}$$

where f denotes the conditional normal density function for Y_1 or Y_2 given d_i. The conditional densities are functions of the parameters β_{1k}, β_{2k}, π_k, σ_1, σ_2, σ_{13}, and σ_{23}. Maddala (1983:284) provides explicit formulas for the densities.

Maximum likelihood estimation consists of employing a numerical procedure to pick values of the parameters that make L as large as possible. Estimates can be obtained using the program HOTZTRAN (Avery and Hotz, 1983), which is used in the numerical example discused below. As for maximum likelihood estimation generally, one can assess the *relative* fit of two models when one of the two models contains all of the parameters contained in the other model plus additional parameters. If L_1 denotes the likelihood of the simpler model and L_2 denotes the likelihood of the model in which the simpler model is nested, then the statistic $\chi^2 = -2\log(L_1/L_2)$ follows an asymptotic chi-square distribution with degrees of freedom equal to the difference in the number of estimated parameters between the two models. Likelihood ratio tests enable one to test simple restrictions that sets of parameters are insignificant or the more complex restrictions implied by the Maximization and Quota Models. Maximum likelihood estimation also yields an estimate of the asymptotic standard errors of parameters. Ratios of estimated parameters to their standard errors follow an asymptotic standard normal distribution and thus can be interpreted as Z-scores.

Estimation of Restricted Models

The Maximization Model is estimated by imposing nonliner restrictions on the general model. Although it can be estimated by either multi-stage or maximum likelihood methods, the latter are greatly preferred. When more than a single regressor appears in equations (2) and (3), the model is overidentified. Multi-stage estimation provides no easy way to impose all of the overidentifying restrictions simultaneously. Typically, therefore, multiple solutions are available for the model. To yield a unique best solution for the model it is necessary to impose all of the restrictions simultaneously, which is straightforwardly carried out by maximum likelihood.

The Quota Model is also a restricted form of the general model and can, in principle, be estimated as a three-equation model. The trivariate normal distribution of the disturbances, however, reduces to bivariate normal because the latent selection variable Z and the expected outcome in position 1, Y_1, are perfectly correlated. One can, therefore, reformulate the model as a *two*-equation model; that is, a censored regression (tobit) equation (e.g., Maddala, 1983:151-161; Tobin, 1958) with unknown censoring point for Y_1 and a linear regression equation subject to selection for Y_2 with disturbances correlated to estimate σ_{23}. Another modification is to assume a known censoring point, and estimate the model as a tobit equation for Y_1 and a linear regression subject to selection for Y_2.

EMPIRICAL EXAMPLE: ACADEMIC TRACKING AND ACHIEVEMENT

We illustrate endogenous switching models with analyses of the causes and effects of allocating students to college and noncollege tracks within secondary schools. We model the processes by which students are assigned to tracks and tracks affects mathematics achievement. Gamoran and Mare, (1987) discuss this research in more detail.

Three Views of Tracking

A rationale for tracking is that students differ in their academic goals and in the environments where they learn best. Ideally, a system of tracking matches students' aptitudes to learning environments. According to this view, student achievement is higher in a tracked high school than in one where tracking is nonexistent or in one where tracks exist but student aptitudes are not matched to track programs. A further goal of tracking is to raise the achievement of students of low academic promise and thus to reduce inequalities in achievement among students.

Critics of academic tracking stress the potential of tracking systems to widen differences between students. Tracks produce larger academic and post-schooling inequalities among students than would exist in the absence of tracking. Moreover, because of the potential stigma and uneven quality of instruction attached to non-college tracks, some students may learn *less* or be less likely to realize their academic or vocational goals when assigned to a non-college track than they could in a different track or than they could achieve in an untracked high school system.

A third view is that tracking is neutral. That is, tracking systems affect neither inequalities among groups of students nor students' average levels of achievement, competency, or post-high school success.

Many quantitative studies of tracking incorporate the effects of tracking into linear ANACOVA models of academic achievement and social stratification. Typically, such studies classify students by track and treat track as a variable that "intervenes" between family, school, and early achievement on the one hand and later achievement on the other. While informative, these analyses are ill-suited to show whether: (1) apparent achievement differences between tracks are due to participation in the tracks themselves or to unmeasured differences between students that are correlated with track placement; (2) the system of tracking widens or narrows achievement differences between groups that would occur under alternative track assignments or in the absence of tracking; and (3) students are optimally (for their aptitudes) assigned to tracks or could potentially perform better if they were assigned to different tracks. These questions are at the heart of an appraisal of the value and cost of tracking.

Models of Tracking Effects

We apply the models discussed above to the study of track placement and track effects in order to answer important analytic questions more fully than in past research. These models enable one to estimate the effects of school, family, and ability factors on track placement and on schooling outcomes, taking into account that students are assigned to tracks on a nonrandom basis. According to the models, students are systematically selected into tracks on the basis of their known characteristics and the largely unobservable beliefs of teachers, administrators, parents, and students themselves about their "suitability" to a particular track. Both measured and unmeasured factors have *common* effects on both achievement and on track placement. Effectively, therefore, these models control for unmeasured factors that potentially distort estimates of track effects in models used in previous work on tracking. These models also allow for *distinct* latent variables to affect achievement in each track. The unmeasured aptitudes that may govern track placement and subsequent achievements may differ across tracks. This allows for explicit estimation and testing of the presumed beneficial or harmful effects of tracking discussed above.

Data and Measures

The data for this analysis are from the High School and Beyond Survey (HSB) of students who were sophomores in 1980. Family, school, and achievement measures were obtained in this survey for sophomores and achievement measures were obtained again in 1982 when most of these students were seniors (Jones *et al.*, 1983). Our analyses are for public school students who were interviewed and tested in both waves of the survey, including dropouts, but excluding students who graduated or transferred before the 1982 interview. We use the following measures: mathematics achievement scores in 1980 and 1982; sophomore achievement scores on science, reading, vocabulary, writing, and civics; socioeconomic status, which is an unweighted linear composite of father's occupational prestige score, father's and mother's grades of school completed, family income, and a home artifacts scale; dichotomous variables that equal one for females, blacks, and Hispanics, and zero otherwise; mean socioeconomic status of school attended; percent black in school attended; percent Hispanic in school attended; number of advanced mathematics courses offered in school attended; and a dichotomous variable that equals one if the student reported being in a college preparatory (academic) track in sophomore year and zero otherwise. Gamoran (1986) and Jones *et al.* describe these measures in more detail. We retained only students for whom data are present for all variables listed above. To speed computations, we used a 25 percent random subsample of observations, resulting in a sample of 3377 persons. Table 1 reports means and standard deviations of the variables for the total and track-specific samples.

Empirical Models

We formulate models for senior year mathematics achievement. The models include equations for: (1) mathematics achievement within the college-track, (2) mathematics achievement within the noncollege track, and (3) the probability of assignment to the college track in sophomore year. These models represent the effects of family, school, and achievement factors as of sophomore year on track placement and subsequent achievement. The models allow common unmeasured variables to affect both track placement and achievement in a track.

In the *structural* versions of the models, senior mathematics achievement is a function of track placement and sophomore achievement in mathematics, science, reading, vocabulary, writing, and civics; and track placement is a function of expected achievement in college and noncollege tracks, sophomore achievement, and sociodemographic and school variables listed above. This model corresponds to equations (2), (3), and (4), where Y_1 and Y_2 denote senior achievement in the college and noncollege tracks respectively; the X_k in equations (2) and (3) denote sophomore achievement scores and, in equation (4), denote sociodemographic and school factors as well as sophomore achievement; and Z indexes the chances of assignment to the college track. Sociodemographic and school factors do not affect achievement directly, but these variables do affect achievement through their effects on both sophomore achievement and on track assignment.

We estimate several versions of this model, including (1) a general model that allows for nonzero covariances between the reduced form disturbances (σ_{13} and σ_{23}) and no restrictions on the slope parameters; (2) an Ascription Model that does not restrict the slope parameters, but constrains σ_{13} and σ_{23} to be zero; (3) a Maximization Model that constrains both slopes and σ_{13} and σ_{23}; (4) a Quota model that imposes alternative slope and covariance restrictions. In these examples the slope parameters are free to vary across the college and noncollege track equations for mathematics achievement. All models are estimated by maximum likelihood, implemented by HOTZTRAN.

Table 2 presents -2 log likelihood statistics and numbers of parameters for each of the four models. The absolute values of these statistics have no interpretation, but their relative values indicate relative fit. Models 2, 3, and 4 are nested within Model 1, but none of the restricted models is nested within any of the other two. The χ^2 statistics show that the general model fits significantly better than both the Quota and Maximization Models, implying that the restrictions on the latter two models are not consistent with the data. In contrast, the Ascription Model, which assumes that track assignment is based on measured characteristics alone, fits as well as the general model. These results suggest that simple views of how students are allocated to tracks--such as that students are matched to the tracks where they are expected to perform best; or that a "quota" of the most promising students is chosen for the college track--are not adequate to describe the data. They also suggest that, at least for the specification used here, common, unmeasured determinants of track placement and achievement do not seriously distort estimates of the effects of measured variables.

The Effects of Measured Variables

Although the general model with correlated disturbances does not fit the data significantly better than the Ascription model, we discuss the general model in further detail to illustrate the interpretation of the disturbance covariances. Table 3 reports the *reduced form* parameter estimates for the general model.

The slope coefficients mainly reflect well-known effects of social factors on achievement and track placement. Sophomore achievement in all subjects except civics affects senior mathematics achievement, and sophomore mathematics achievement has by far the strongest effect. These factors explain approximately 73 and 66 percent of the variance in mathematics achievement in the college and noncollege tracks respectively. Except for science, all sophomore achievement scores positively affect the chances of assignment to the college track. Socioeconomic status also positively affects college track placement. The effects for blacks and Hispanics on track placement suggest that minorities who have *equivalent* sophomore achievement levels to those of white non-Hispanics are more likely to enter the college track. Achievement levels of these groups, however, are significantly lower than average (e.g., Gamoran, 1986) and, as Table 1 shows, Hispanics are substantially *under*represented in the college track.

The Effects of Unmeasured Factors

The disturbance covariances, σ_{13} and σ_{23}, provide information about unobserved influences on tracking and achievement. As Table 2 shows, these parameters are not statistically significant when considered together. Their Z-scores, however, indicate that whereas σ_{13} is insignificant, σ_{23} is significantly less than zero. For the college track, this suggests that students who actually enter the track score no higher on senior mathematics than would a random sample from all sophomores *who have equal sophomore achievement levels* if they were placed in that track. In other words, OLS estimates of the achievement equation for the college track have no selection bias.

For the noncollege track the selection bias is again small, but larger than for the college track. [The estimates for σ_{13} and σ_{23} imply *correlations* of -.051/4.094 = -.012 and -.334/4.343 = -.077 respectively.] That σ_{23} is negative indicates that students who in fact enter the noncollege track do somewhat *better* on senior mathematics than would a random sample of all sophomores *who have equal sophomore achievement levels* if they were placed in that track. Selection into the noncollege track, therefore, is biased in favor of students who can perform well in that track compared to other students with the same attributes. Thus OLS potentially *overestimates* achievement within the noncollege track because it ignores positive selection into that track.

The Effects of Tracking on Achievement

The results discussed thus far suggest that observed net differences in achievement between tracks understate actual differences due to tracks because of the positive selection of students into the noncollege track. It remains to be shown, however, whether students who enter the noncollege track score higher on senior mathematics than they would were they assigned to the *college* track.

To investigate tracking effects more fully we calculate expected levels of achievement that students would attain under alternative track assignments. We use the parameter estimates for college and noncollege achievement reported in Table 3 and apply equations (6)-(9) for a hypothetical individual who has the average values of the total sample on the sophomore achievement tests and a probability of assignment to the college track equal to the sample proportion of persons in the college track (see Table 1). Table 4 summarizes the calculations. The columns of the table compare tracks and the rows denote alternate groups of persons: actual college track students, actual noncollege track students, and all students combined.

The final row of the table provides estimates of achievement that would be observed if students were *randomly* assigned to tracks. That is, these estimates are free from selection biases and their difference denotes the structural track effect. The "College-College" and "Noncollege-Noncollege" cells of the table denote the conditions that the students actually experienced. The "College-Noncollege" and "Noncollege-College" cells denote hypothetical levels of achievement that students would experience were they assigned to a different track from the one that they in fact experienced.

All the contrasts in Table 4 are very small--a fraction of a standard deviation of senior mathematics achievement and much smaller than the observed difference between tracks on senior achievement (see Table 1). This, of course, results from our adjustment of all effects to the sample averages for sophomore level achievement. The following discussion, therefore, is useful for illustrating the model, but it is based on contrasts of limited substantive importance. The main results in Table 4 are as follows:

(1) The college track produces higher mathematics achievement than the noncollege track for the population as a whole and, hypothetically, within the populations defined by persons who in fact entered one track or the other. For the total population, the advantage of the college track is 20.86 - 19.59 = 1.27, an estimate of the difference in achievement that would be observed were students with equal sophomore achievement levels randomly assigned to tracks.

(2) This effect of tracking is *larger* than the actual contrast between college and noncollege students (20.80 - 19.76 = 1.04). Relative to a system in which students are randomly assigned to tracks, the actual system has smaller achievement differentials between the tracks. This results because noncollege students are positively selected into the noncollege track, but selectivity is negligible for the college track.

(3) The existing tracking regime slightly increases the average *level* of achievement in the population. Assuming that the proportion of students in the college track is fixed at .32, actual achievement is 20.09 [that is, (.32)(20.80) + (.68)(19.76)]; whereas under random assignment it is 20.00 [that is, (.32)(20.86) + (.68)(19.59)]. This is also a result of positive selection into the noncollege track.

(4) Paradoxically, the actual noncollege track students would do at least as well as the actual college track students if they competed in the same track. *Within* both college and noncollege tracks, the actual noncollege

students have slightly higher levels of achievement (a difference of 20.88 - 20.80 = 0.08 in the college track and of 19.76 - 19.22 = 0.54 in the noncollege track). Because the *between* track difference is so strong, however, the noncollege track students do not actually fare as well as the college track students.

(5) Although the relative standing of the noncollege students is most favorable in the noncollege track (and advantage of 0.54 in the noncollege track versus 0.08 in the college track), their absolute standing is greatest in the college track (20.88 versus 19.76).

In summary, the existing system of tracking narrows the achievement gap between the actual college and noncollege students and produces higher achievement than under random tracking; but, noncollege students could increase their achievement if more of them enrolled in the college track.

The results presented have been simplified for didactic purposes. A fuller study of tracking effects requires the investigation of other models and outcomes of tracking. An important extension of our models is the analysis of more than one outcome. A regime of tracking may, for example, have different effects on mathematics or other types of achievement than on rates of high school graduation. Decision makers may explicitly choose a track for students that does not maximize their achievement but rather yields an optimal combination of expected achievement and expected graduation probabilities. The combined effects of tracking on achievement and high school graduation can be assessed using extensions of the models discussed here.

CONCLUSION

Sociological investigations commonly assume that whereas the structure of social positions is relatively immune to the types of individuals that occupy the positions, positions have substantial effects on individuals. Although this assumption is often valid, many investigators erroneously infer that the characteristics of positions are exogenous variables in empirical models that predict individual outcomes. When social actors have discretion over how persons and positions are matched, then the characteristics of the positions and of the individuals that occupy them are *jointly* determined.

This chapter has described models that can be used to study the consequences of social positions when both allocation to the positions and their consequences have common causes. These models are likely to have their greatest value when investigators wish to assess explicit hypotheses about the linkage of social positions and their occupants. When the goals of the analyst are more exploratory and descriptive, then models such as ANACOVA, which make simpler assumptions about unobserved variables, may be more useful.

REFERENCES

Aldrich, John H., and Forrest D. Nelson. 1984. *Linear Probability, Logit, and Probit Models*. Quantitative Applications in the Social Sciences No. 45. Beverly Hills, CA: Sage University Press.

Amemiya, T. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Avery, R. B., and V. J. Hotz. 1983. "HOTZTRAN User's Manual." Unpublished Manuscript. Economics Research Center/NORC. Chicago.

Berk, R. A. 1983. "An Introduction to Sample Selection Bias in Sociological Data." *American Sociological Review* 48:386-398.

Gamoran, A. 1986. "The Stratification of High School Learning Opportunities." Paper presented to the Annual Meetings of the American Educational Research Association, San Francisco.

Gamoran, A., and R. D. Mare. 1987. "Secondary School Tracking and Stratification: Compensation, Reinforcement, or Neutrality?" Paper presented to the Annual Meetings of the American Educational Research Association, Washington, D.C.

Garen, J. 1984. "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable." *Econometrica* 52:1199-1218.

Goldfeld, S. M., and R. E. Quandt. 1973. "The Estimation of Structural Shifts by Switching Regressions." *Annals of Economic and Social Measurement* 2:475-86.

Heckman, J. J. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46:931-960.

Heckman, J. J., and R. Robb Jr. 1985. "Alternative Methods for Evaluating the Impact of Interventions." Pp. 156-245 in J. J. Heckman and B. Singer (eds.), *Longitudinal Analysis of Labor Market Data*. Cambridge: Cambridge University Press.

Heckman, James J., and R. Robb Jr. 1986. "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes." Economics Research Center Discussion Paper, 86-9. NORC. Chicago.

Johnston J. 1984. *Econometric Methods*, Third Edition. New York: McGraw-Hill Book Company.

Jones, C., M. Clark, G. Mooney, H. McWilliams, I. Crawford, B. Stephenson, and R. Tourangeaou. 1983. *High School and Beyond 1980 Sophomore Cohort First Follow-Up (1982) Data File User's Manual*. Washington, D.C.: National Center for Education Statistics.

Long, J. S., and T. D. Miethe. 1987. "The Statistical Comparison of Groups." In. J. S. Long (ed.), *Common Problems in Quantitative Social Research*. Beverly Hills, CA: Sage University Press.

Maddala, G.S. 1977. *Econometrics*. New York: McGraw-Hill Book Company.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Tobin, J. 1958. "Estimation of Relationships for Limited Dependent Variables." *Econometrica* 26:24-36.

Willis, R. J., and S. Rosen. 1979. "Education and Self-Selection." *Journal of Political Economy* 87, Part 2:S7-S36.

Winship, C., and R. D. Mare. 1983. "Structural Equations and Path Analysis for Discrete Data." *American Journal of Sociology* 89:54-110.

Winship, C., and R. D. Mare. 1984. "Regression Models with Ordinal Variables." *American Sociological Review* 49:512-525.

TABLE 1. MEANS AND STANDARD DEVIATIONS OF VARIABLES IN TRACKING ANALYSIS FOR TOTAL, AND TRACK-SPECIFIC SAMPLES*

Variable	Sample					
	Total		College Track		Noncollege Track	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Senior Math Achievement	19.97	8.22	24.72	7.81	17.72	7.42
College Track (vs. Other)	0.32	0.47	1.00	0.00	0.00	0.00
Sophomore Achievement						
Mathematics	18.79	7.41	22.82	7.16	16.88	6.72
Science	11.13	3.74	12.63	3.59	10.42	3.59
Reading	9.11	3.92	11.01	3.94	8.21	3.57
Vocabulary	10.97	4.32	12.95	4.15	10.03	4.08
Writing	10.29	3.92	12.20	3.41	9.38	3.92
Civics	5.88	2.04	6.67	1.89	5.50	2.00
Female (vs. Male)	0.48	0.50	0.53	0.50	0.46	0.50
Socioeconomic Status	-0.09	0.73	0.19	0.73	-0.22	0.69
Black (vs. Nonblack)	0.12	0.32	0.11	0.31	0.12	0.33
Hispanic (vs. Non-Hisp.)	0.13	0.34	0.08	0.28	0.15	0.36
School SES	-0.09	0.36	-0.01	0.37	-0.12	0.35
School % Black	13.34	22.13	12.72	21.33	13.63	22.49
School % Hispanic	4.57	25.12	4.26	11.51	4.72	12.90
School Math Courses	3.42	0.73	3.53	0.67	3.37	0.75

*Observations are weighted. See text for discussion of variables and samples.

TABLE 2. LIKELIHOOD STATISTICS AND χ^2 TESTS FOR SELECTED MODELS OF MATHEMATICS ACHIEVEMENT AND TRACK ASSIGNMENT

Model	General	Ascription	Maximization	Quota
-2 Log Likelihood	46074	46076	46724	46180
Number of Parameters	33	31	26	26
Likelihood Ratio Test vs. General Model				
χ^2		2	650	1062
D.F.		2	7	7
p		$.3 < p < .5$	$p < .001$	$p < .001$

TABLE 3. REDUCED FORM PARAMETERS FOR MODEL OF TRACK ASSIGNMENT AND MATHEMATICS ACHIEVEMENT IN SENIOR YEAR

Independent Variable	Dependent Variable: College Track Assignment		Achievement in College Track		Achievement in Noncollege Track	
	π	Z*	β	Z*	β	Z*
Constant	-2.815	-22.7	1.355	3.5	0.816	3.5
Sophomore Achievement						
Math	0.040	11.0	0.725	39.6	0.709	49.5
Science	-0.001	-0.2	0.177	4.8	0.220	8.3
Reading	0.024	3.3	0.130	3.5	0.055	2.0
Vocabulary	0.020	3.2	0.079	2.5	0.103	4.4
Writing	0.033	4.7	0.137	3.9	0.139	5.8
Civics	0.029	2.6	0.076	1.3	-0.010	-0.2
Female	0.141	3.7				
SES	0.269	9.3				
Black	0.281	4.1				
Hispanic	0.078	1.3				
School SES†	-0.103	-1.7				
% Black†	0.006	6.0				
% Hispanic†	0.004	3.5				
Math Courses†	0.124	4.9				
R^2			0.725		0.657	
σ	1.000		4.094	68.2	4.343	94.3
σ_{13}, σ_{23}			-0.051	-0.8	-0.334	-4.0

† Characteristics of high school that respondent attended in 1980.
* Z denotes ratio of estimated coefficient to its asymptotic standard error.

TABLE 4. SUMMARY OF HYPOTHETICAL EFFECTS OF ASSIGNMENT TO ALTERNATIVE TRACKS ON SENIOR MATHEMATICS ACHIEVEMENT*

	Track	
	College	Noncollege
Population		
College	20.80	19.22
Noncollege	20.88	19.76
All	20.86	19.59

* Effects are estimated holding constant levels of sophomore achievement on all academic subjects at total sample means and assuming a probability of assignment to college track of .32.

Mailing Address:

Center for Demography and Ecology
University of Wisconsin
1180 Observatory Drive
Madison, Wisconsin 53706-1393
U.S.A.