Annu. Rev. Sociol. 1992. 18:327–50

MODELS FOR SAMPLE SELECTION BIAS

Christopher Winship

Department of Sociology, Northwestern University, 1810 Chicago Avenue, Evanston, Illinois 60201

Robert D. Mare

Department of Sociology, University of Wisconsin, 1180 Observatory Drive, Madison, Wisconsin 53706

KEY WORDS: sampling, selection bias, methodology, statistics

Abstract

When observations in social research are selected so that they are not independent of the outcome variables in a study, sample selection leads to biased inferences about social processes. Nonrandom selection is both a source of bias in empirical research and a fundamental aspect of many social processes. This chapter reviews models that attempt to take account of sample selection and their applications in research on labor markets, schooling, legal processes, social mobility, and social networks. Variants of these models apply to outcome variables that are censored or truncated—whether explicitly or incidentally—and include the tobit model, the standard selection model, models for treatment effects in quasi-experimental designs, and endogenous switching models. Heckman's two-stage estimator is the most widely used approach to selection bias, but its results may be sensitive to violations of its assumptions about the way that selection occurs. Recent econometric research has developed a wide variety of promising approaches to selection bias that rely on considerably weaker assumptions. These include a number of semi- and nonparametric approaches to estimating selection models, the use of

0732-0582/92/0410-0327$02.00

panel data, and the analyses of bounds of estimates. The large number of available methods and the difficulty of modelling selection indicate that researchers should be explicit about the assumptions behind their methods and should present results that derive from a variety of methods.

INTRODUCTION

Sample selection is a generic problem in social research that arises when an investigator does not observe a random sample of a population of interest. Specifically, when observations are selected so that they are not independent of the outcome variables in the study, this sample selection leads to biased inferences about social processes. A wide variety—perhaps the majority—of research traditions in sociology rely on designs that are vulnerable to sample selection biases. To rely exclusively on observational schemes that are free from selection bias is to rule out a vast portion of fruitful social research. Indeed, to understand how social positions affect the behaviors of their incumbents, one often must study the processes through which individuals are selected into such positions. Selectivity is not only a source of bias in research, but also the subject of substantive research.

An intuitive appreciation of the ways that selection bias affects inference has always been part of sound research practice. In recent decades, however, many social scientists have formalized the ways that selectivity can affect inferences about social processes through the use of *models* for sample selection bias. These models demonstrate formally how and why bias comes about, and they also show the common formal structure of an array of substantive investigations affected by sample selection bias.

In a linear regression model, selection occurs when data on the dependent variable are missing nonrandomly conditional on the independent variables. Elementary statistical methods in this situation generally yield biased and inconsistent estimates of the effects of the independent variables. For example, if a researcher uses ordinary least squares (OLS) to estimate a regression model where large values of the dependent variable are underrepresented in a sample, the estimates of slope coefficients may be biased.

Sociologists increasingly use models to take account of sample selection bias. A growing methodological literature has also focused on the general issue of the contaminating influence of nonrandom selection on causal inference (Berk 1988, Lieberson 1985). Outside of sociology, especially in economics, applied and theoretical research on selection has been much more extensive, yielding many hundreds of articles. The recent literature on models for sample selection bias develops three major themes: (*a*) Selection is pervasive and results naturally from human behavior (e.g. Roy 1951, Gronau 1974, Heckman 1974, Lewis 1974, Willis & Rosen 1979, Heckman &

Sedlacek 1985, 1990, Heckman & Honore 1990); (*b*) models for sample selection share a close affinity with models for assessing program treatment and other types of effects in experimental and nonexperimental contexts (e.g. Ashenfelter 1978, Barnow et al 1980, Lalonde 1986, Heckman & Robb 1985, 1986a,b, Heckman & Hotz 1989); (*c*) models for selection bias are only as good as their assumptions about the way that selection occurs, and estimation strategies are needed that are robust under a variety of assumptions (Arabmazar & Schmidt 1982, Goldberger 1983, Lee 1982, Wainer 1986, Barnett et al 1991).

This article reviews the significance of selection bias in social research, the problem of modeling selection, and technical issues that arise in correcting for selection bias; we emphasize recent econometric research. We focus on the problem of estimating a linear regression model in the presence of selection. Because Heckman's (1979) estimator has been used extensively in the recent social science literature, we emphasize its problems and extensions. In this review we: (*a*) show why selection on the dependent variable leads to biased and inconsistent estimates of parameters in a regression model; (*b*) review contexts in which selection arises in sociological research and consider some of the models that have been proposed; (*c*) provide a simple classification of alternative selection models; (*d*) discuss Heckman's estimator and its limitations; (*e*) describe semiparametric and nonparametric generalizations of Heckman's estimator; and (*f*) discuss other approaches to selection, including Manski's bound approach and methods that rely on panel data.

An exhaustive review of the literature on issues related to selectivity is impossible within the available space. We emphasize material that is unfamiliar to most sociologists at the neglect of other topics. Berk (1983) and Berk & Ray (1982) introduce selection models to sociologists, and Maddala (1983) and Amemiya (1985) summarize the literature developed during the 1970s that is concerned with estimators other than Heckman's. We do not discuss the statistics literature on missing data (Little & Rubin 1987), and we touch only briefly on causal inference in nonexperimental research and the closely related issue of social program evaluation. These issues have spawned a substantial recent literature (e.g. Lieberson 1985, Holland 1986, Wainer 1986, Berk 1988, Marini & Singer 1988, Manski & Garfinkel 1992).

THE STRUCTURE OF SELECTION

We illustrate selection bias for a single regression equation. The ideas presented here extend easily to more complex models, including those with discrete and other limited-dependent variables and those with multiple dependent variables in the regression model. Selection bias results from a correlation between the error and the independent variables. Consider an example, first used by Hausman & Wise (1977) in their discussion of selection bias. In the 1970s the US government supported several income maintenance experiments, which were based on samples of families with incomes below a specified level. The experiments were designed to reveal whether income supplements for poor persons affect their willingness to work, but the data from the experiments have proved useful for other investigations as well (e.g. Hannan et al 1977, 1978). In the following discussion, we ignore the original purpose of the experiments and focus on problems created when the control subjects of the experiments are used to answer other research questions.

Truncated Samples—Explicit Selection

Consider the problem of estimating the effect of education on income from a sample of persons with incomes below $15,000. This is shown in Figure 1, where individuals are sampled at three education levels: low (L), medium (M), and high (H). When observations with values of the dependent variable that are beyond a certain bound are excluded, the resulting sample is *truncated.* This is also termed explicit selection inasmuch as whether an observation enters the sample is an exact function of the dependent variable (Goldberger 1981). In Figure 1, sample truncation leads to an estimate of the effect of schooling that is biased downward from the true regression line, a result of the $15,000 ceiling on the dependent variable. Under certain conditions—that is, if there is only a single regressor, if the distribution of the independent variables is multivariate normal, or if the independent variables follow a specific class of stable distributions—then all the regression coefficients are biased downwards (Goldberger 1981, Ruud 1986). In general, however, selection may bias estimated effects in either direction.

A sample that is restricted on the dependent variable is effectively selected on the error of the regression equation; at any value of the independent variables, observations with sufficiently large positive errors are eliminated from the sample. As shown in Figure 1, as the independent variable increases, the expected value of the error becomes increasingly negative, making them negatively correlated. Because this contradicts the standard assumption of OLS that the error and the independent variables are uncorrelated, OLS estimates are biased.

Censored Samples—Explicit Selection

A different type of explicit selection occurs when the sample *includes* persons with incomes of $15,000 or more, but all that is known about such persons is their educational attainment and that their incomes are $15,000 or greater. When the dependent variable is outside of a known bound but the exact value of the variable is unknown, the sample is *censored.* If these persons' incomes

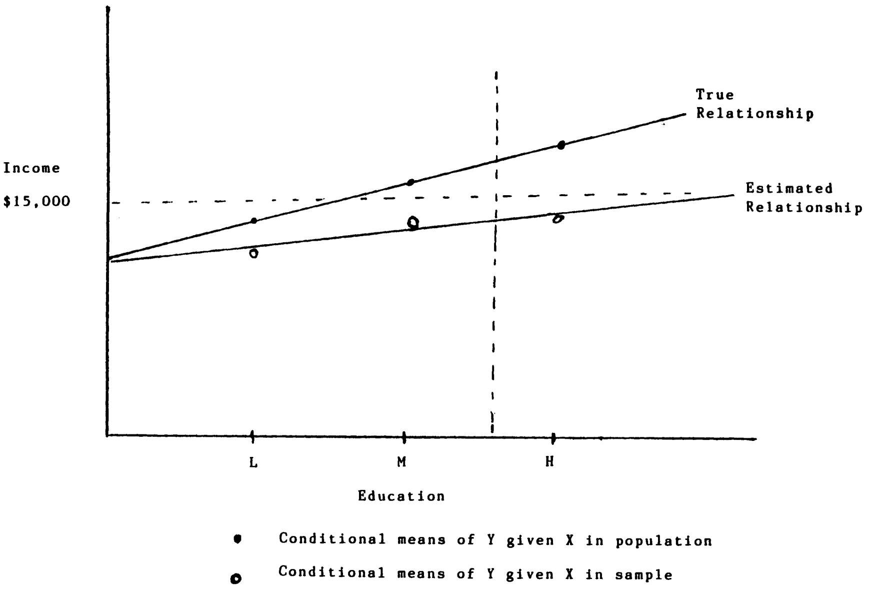

Figure 1 Estimating the effect of education on income from a sample of persons with incomes below $15,000. Samples are at three education levels: low (L), medium (M), and high (H).

are coded as $15,000, OLS estimates are biased and inconsistent for the same reasons as in the truncated sample. Since true incomes are unknown, the expected value of the error at any level of education is negative and becomes increasingly negative as education increases, in contradiction to the OLS assumption that the expected value of the error is zero.

Censored and Truncated Samples—Incidental Selection

A third type of selection occurs when censoring or truncation is a stochastic function of the dependent variable. In the example, the *probability* that income is unobserved is a function of income or, equivalently, a function of education and the error. This is termed incidental selection (Goldberger 1981). As we show below, biases in OLS estimates similar to those for explicit selection are the result.

Selection on Measured Independent Variables

Yet another type of selection occurs when the dependent variable is missing solely as a function of the measured independent variable(s); for example, the sample is selected on educational attainment alone. If persons with high levels of schooling are omitted from the model, an OLS estimate of the effect on income for persons with lower levels of education is unbiased if schooling has a constant linear effect throughout its range. Because the conditional expectation of the dependent variable at each level of the independent variable is unaffected by a sample restriction on the independent variable, when a model is properly specified OLS estimates are unbiased and consistent (DuMouchel & Duncan 1983).

SOCIAL SCIENCE EXAMPLES OF SELECTION BIAS

In this section we review examples drawn from the sociology literature of the effects of sample selection bias and of approaches designed to take selectivity into account. Many of these examples illustrate where it has been fruitful both to correct for selection bias and to incorporate the selection process into the substantive investigation. In econometrics, where much basic research on selection bias has been done, many of the applications have been to labor economics. Many studies by sociologists that deal with selection problems have been in the cognate area of social stratification. Problems of selection bias, however, pervade sociology, and attempts to grapple with them appear in the sociology of education, family sociology, criminology, the sociology of law, social networks, and other areas. We select examples where analysts have used models for selection bias, but one could name many other cases where selection biases exist but have thus far been neglected.

Trends in Employment of Out-of-School Youths

Mare & Winship (1984) investigate employment trends from the 1960s to the 1980s for young black and white men who are out of school. Many factors affect these trends, but a key problem in interpreting the trends is that they are influenced by the selectivity characteristic of the out-of-school population. Over time, the selectivity changes because the proportion of the population that is out of school decreases, especially among blacks. Because persons who stay in school longer have better average employment prospects than do persons who drop out, the employment rates of nonstudents are lower than they would be if employment and school enrollment were independent (Mare et al 1984). Observed employment patterns are biased because the probabilities of employment and leaving school are dependent. *Ceteris paribus,* as enrollment increases, employment rates for out-of-school young persons decrease. To understand the employment trends of out-of-school persons, therefore, one must analyze jointly the trends in employment and school enrollment. The increasing propensity of young blacks to remain in school explains some of the growing gap in the employment rates between blacks and whites (Mare & Winship 1984). In this case selectivity is a key part of the substantive interpretation.

Selection Bias and the Disposition of Criminal Cases

A central focus in the analysis of crime and punishment are the determinants of differences in the treatment of persons in contact with the criminal justice system; for example, the differential severity of punishment of blacks and whites (Peterson & Hagan 1984). A sample of persons who are convicted of crimes is highly selective. Of those who commit crimes only a portion are arrested; of those arrested, only a portion are prosecuted; of those prosecuted, only a portion are convicted; of those convicted, only a portion are sent to prison. Common unobserved factors may affect continuation from one stage of this process to the next. Indeed, the stages may be jointly determined inasmuch as legal officials may process cases mindful of the likely outcomes later in the process. The chances that a person will be punished if arrested, for example, may affect the eagerness of police to arrest suspects. Analyses of the severity of sentencing that focus on persons already convicted of crimes may be subject to selection bias and should take account of the process through which persons are convicted (Hagan & Parker 1985, Peterson & Hagan 1984, Zatz & Hagan 1985).

Scholastic Aptitude Tests and College Success

Manski & Wise (1983) investigate the determinants of graduation from college, including the capacity of Scholastic Aptitude Test (SAT) scores to predict individuals' probabilities of graduation. Studies based on samples of students within colleges find that SAT scores have little predictive power. Yet these studies may be biased because of the selective stages between taking the SAT and attending college. Some students who take the SAT do not apply to college; some apply but are not admitted; some are admitted but do not attend; and those who attend are sorted among the colleges to which they have been admitted. Each stage of selection is nonrandom and is affected by characteristics of students and schools that are unknown to the analyst. When one jointly considers the stages of selection in the college attendance decision, along with the probability that a student graduates from college, one finds that the SAT score is a strong predictor of college graduation.

Women's Socioeconomic Achievement

Analyses of the earnings and other socioeconomic achievements of women are potentially affected by nonrandom selection of women into the labor market. The rewards that women expect from working affect their propensities to enter the labor force. Outcomes such as earnings or occupational status, therefore, are jointly determined with labor force participation, and analyses that ignore the process of labor force participation are potentially subject to selection bias. Many studies in economics (e.g. Gronau 1974, Heckman 1974, 1979) and sociology (Fligstein & Wolf 1978, Hagan 1990, England et al 1988) use models that represent simultaneously women's labor force participation and the market rewards that they receive.

Analysis of Occupational Mobility from Nineteenth Century Censuses

Nineteenth Century Decennial Census data for cities provide a means of comparing nineteenth and twentieth century regimes of occupational mobility in the United States (Grusky 1986, Hardy 1989). Although one can analyze mobility by linking the records of successive censuses, linkage is only possible for persons who remain in the same city and keep the same name over the decade. Persons who die, emigrate, or change their names are excluded. Because mortality and migration covary with socioeconomic success, the process of mobility and the way that observations are selected for the analysis are jointly determined. Analyses that jointly model mobility and sample selection offer the possibility of avoiding selection bias (Hardy 1989).

Bias in Network Analysis

One concern of social network studies is to examine the consequences of social network structure for individuals; for example, Marsden & Hurlbert (1987) examine the effects of network density—that is, the strength of ties that a person has with others with whom they discuss important matters—on personal happiness. Network density is observable only for persons with enough contacts for density measures to be computed; isolates have no network at all. Because isolation is both a cause and a consequence of one's happiness, analyses that exclude isolates are subject to selection bias. Number of contacts and the outcomes of network structure can be analyzed jointly to take the potential selection bias into account.

MODELS OF SELECTION

We now provide a brief classification of selection models at varying levels of complexity. We start by discussing the censored regression or tobit model. Due to limited space we forego discussion of the very closely related truncated regression model (see Hausman & Wise 1976, 1977). For more detailed classifications, see Amemiya (1985) and Heckman (1987).

Tobit Model

The censored regression or tobit model is appropriate when the dependent variable is censored at some upper or lower bound as an artifact of how the data are collected (Tobin 1958, Maddala 1983). For censoring at a lower bound, the model is:

$$Y^*_{1i} = X_i\beta + \epsilon_i \qquad 1.$$

$$Y_{1i} = Y^*_{1i} \text{ if } Y^*_{1i} > 0 \qquad 2.$$

$$Y_{1i} = 0 \text{ if } Y^*_{1i} \leq 0, \qquad 3.$$

where, for the ith observation, Y^*_{1i} is an unobserved continuous latent variable, Y_{1i} is the observed variable, X_i is a vector of values on the independent variables, ϵ_i is the error, and β is a vector of coefficients. We assume that ϵ_i is uncorrelated with X_i and is independently and identically distributed. The model can be generalized by replacing the threshold zero in Equations 2 and 3 with a known nonzero constant. The censoring point may also vary across observations, leading to a model that is formally equivalent to models for survival analysis (Kalbfleisch & Prentice 1980, Lancaster 1990).

OLS estimates of Equation 1 are subject to selection bias. For observations for which $Y_{1i} > 0$, the model implies

$$\begin{aligned} Y_{1i} &= X_i\beta + E[\epsilon_i \mid Y^*_{1i} > 0] + \eta_i \\ &= X_i\beta + E[\epsilon_i \mid \epsilon_i > -X_i\beta] + \eta_i \end{aligned} \qquad 4.$$

where η_i is the difference between ϵ_i and $E[\epsilon_i \mid Y^*_{1i} > 0]$ and is uncorrelated with both terms. Selection bias results because $E[\epsilon_i \mid \epsilon_i > -X_i\beta]$ in Equation 4 is a function of $-X_i\beta$. The less $-X_i\beta$, that is, the less the rate of censoring, the greater is the conditional expected value of ϵ_i. The negative correlation between $-X_i\beta$ and ϵ_i implies that OLS estimates of the regression of Y_i on X_i are biased and inconsistent. An equation analogous to Equation 4 can be constructed for observations for which $Y_{1i} = 0$, producing a parallel analysis. Thus, inclusion of observations for which $Y_{1i} = 0$ leads to similar problems. Equation 4 also shows how selectivity bias may be interpreted as an omitted variable bias (Heckman 1979). The term $E[\epsilon_i \mid Y^*_{1i} > 0]$ can be thought of as an omitted variable that is correlated with X_i and affects Y_1. Its omission leads to biased and inconsistent OLS estimates of β.

Mare & Chen (1986) use the tobit model to examine the effects of parents' socioeconomic characteristics on years of graded schooling completed by their offspring, a variable that is censored for persons with more than 12 years of school. Seltzer & Garfinkel (1990) and Seltzer (1991) use tobit models to analyze the determinants of property and child support awards to mothers and amounts paid by noncustodial fathers after divorce, which are zero for substantial proportions of mothers. In studying how families finance college educations, Steelman & Powell (1989) construct tobit models of the sources of college funding, including parents' contributions, loans, savings, and scholarships, each of which has a logical floor of zero. Hoffman (1984) uses a tobit model to examine the determinants of pious bequests in rural Lyonnais and Beaujolais between 1521 and 1737.

Standard Sample Selection Model

A generalization of the tobit model is to specify that a second variable Y^*_{2i} affects whether Y_{1i} is observed or not. That is, retain the basic model Equation 1, but replace 2 and 3 with:

$$Y_{1i} = Y^*_{1i} \text{ if } Y^*_{2i} > 0 \qquad 5.$$

$$Y_{1i} = 0 \text{ if } Y^*_{2i} \leq 0 \qquad 6.$$

Variants of this model depend on how Y_{2i} is specified. Commonly Y^*_{2i} is determined by a binary regression model:

$$Y^*_{2i} = Z_i\alpha + v_i \qquad 7.$$

$$Y_{2i} = 1 \text{ if } Y^*_{2i} > 0 \qquad 8.$$

$$Y_{2i} = 0 \text{ if } Y^*_{2i} \leq 0, \qquad 9.$$

where Y^*_{2i} is a latent continuous variable. A classic example is a model for the wages and employment of women, where Y_{1i} is the observed wage, Y_{2i} is a dummy variable indicating whether a women works, and Y^*_{2i} indexes a woman's propensity to work (Gronau 1974). In a variant of this model, Y_{2i} is hours of work and Equations 7–9 are a tobit model (Heckman 1974). In both variants, Y^*_{1i} is only observed for women with positive hours of work. One can modify the model by assuming, for example, that Y_{1i} is dichotomous. If ϵ_i and v_i follow a bivariate normal distribution, this leads to a bivariate probit selection model. Maddala (1983) and Lee (1983) discuss these and other variants of the model.

The bias in an OLS regression of Y_{1i} on X_i in the general selection case is similar to that in the tobit model. When $Y^*_{2i} > 0$,

$$\begin{aligned} Y_{1i} &= X_i\beta + E[\epsilon_i \mid Y^*_{2i} > 0] + \eta_i \\ &= X_i\beta + E[\epsilon_i \mid v_i - Z_i\alpha > 0] + \eta_i \end{aligned} \qquad 10.$$

The OLS regression of Y_{1i} on X_i is biased and inconsistent if ϵ_i is correlated with $v_i - Z_i\alpha$, which occurs if ϵ_i is correlated with either v_i or the Z_i. If the variables in Z_i are included in X_i, ϵ_i and Z_i are uncorrelated by assumption. If, however, Z_i contains additional variables, then ϵ_i and Z_i may be correlated.

When $\sigma_{\epsilon v} = 0$ selection depends only on the observed variables in Z_i not in X_i. In this case, propensity score methods (Rosenbaum & Rubin 1983) are appropriate for correcting for selectivity (Heckman & Robb 1985, 1986a,b).

Treatment Effects

The problem of selection is closely related to the problem of estimating treatment effects in the presence of nonrandom assignment. Consider the following model for the effect of a dichotomous variable Y_2 on a continuous variable Y_1:

$$Y_{1i} = X_i\beta + Y_{2i}\gamma + \epsilon_i, \qquad 11.$$

where γ is the treatment effect to be estimated and all other notation is defined as above. We can estimate Equation 11 consistently by OLS if X_i and Y_{2i} are uncorrelated with the error ϵ_i, a condition that is met if assignment to the two treatment levels is random, or random conditional upon the X_i. In the latter case OLS corrects for the correlation between Y_{2i} and X_i in estimating γ. When assignment to Y_{2i} is a function of the error, Y_{2i} is determined endogenously. In this case, methods used to correct for selection can also be used to correct for the endogeneity of Y_{2i} (Heckman, 1976b, 1978). Alternatively, instrumental variable methods can be used if instruments for Y_{2i} are available.

The relationship between the treatment and the selection models can be understood by first considering Rubin & Holland's structure for measuring causal effects (Holland 1986, Rubin 1978). Assume that associated with each observation there are two variables, Y^0_{1i}, which is the outcome on variable Y_1 for observation i when it is assigned to treatment level $Y_{2i} = 0$, and Y^1_{1i} which is the outcome on variable Y_1 for observation i when it is assigned to treatment level $Y_{2i} = 1$. Rubin & Holland then define the causal effect of the treatment for the ith observation as the difference: $Y^1_{1i} - Y^0_{1i}$. The average causal effect, γ, is then the average of this difference across observations. In almost all situations we only observe either Y^0_{1i} or Y^1_{1i} for any given observation. As a result the observation-level and the average treatment effect cannot be directly estimated.

This framework can be generalized to the regression case by rewriting Equation 11 as two equations, one for each value of Y_{2i}:

$$Y^0_{1i} = X_i\beta + \epsilon_{1i} \qquad Y_{2i} = 0 \qquad 12.$$

$$Y^1_{1i} = X_i\beta + \gamma + \epsilon_{2i} \qquad Y_{2i} = 1 \qquad 13.$$

In Equation 13 γ denotes how the intercept differs between when $Y_{2i} = 0$ and when $Y_{2i} = 1$ and is equal to the average treatment effect. By itself Equation

12 is subject to selection bias in that data are "missing" on Y_{1i}^0 when $Y_{2i} = 1$. Likewise Equation 13 is subject to selection bias in that data are "missing" on Y_{1i}^1 when $Y_{2i} = 0$. Unless Y_{2i} is determined randomly (given X_i), OLS estimates, whether taken separately from Equations 12 and 13 or jointly from 11, are subject to selection bias.

The problem of estimating treatment effects is an example of the general problem of causal analysis with nonexperimental data. The assessment of a treatment effect is a "missing data" problem in that, for each case, we observe the dependent variable under only one condition, and the effect of treatment is the difference for the case between the dependent variable under that condition and the alternative condition (Holland 1986, Rubin 1978). Thus, any type of causal analysis is potentially a problem in selection.

Endogenous Switching Regressions

The treatment model can be generalized to the endogenous switching model, which allows the effects of the independent variable to vary across treatments (Maddala 1983, Mare & Winship 1988). Then the model becomes:

$$Y_{1i}^0 = X_i\beta_1 + \epsilon_{1i} \qquad (Y_{2i} = 0) \qquad 14.$$

$$Y_{1i}^1 = X_i\beta_2 + \epsilon_{2i} \qquad (Y_{2i} = 1), \qquad 15.$$

where γ becomes part of the intercept in β_2 and Y_2 is determined by Equations 7–9. This model is suitable for assessing the effects of a social classification Y_2 on a consequence of membership in this classification Y_1—such as the effect of academic track placement on achievement or the effect of labor market sector on earnings—and how the effects of exogenous characteristics vary across levels of Y_2. By itself Equation 14 is subject to selection bias in that data are "missing" on Y_{1i}^0 when $Y_{2i} = 1$. Likewise Equation 15 is subject to bias in that data are "missing" on Y_{1i}^1 when $Y_{2i} = 0$. Only if, conditional on the X_i, observations enter levels of Y_{2i} at random, are OLS estimates of Equations 14 or 15 unbiased.

The covariances of the disturbances in this model provide information about the nature of selectivity into each group. Denote the covariances of Equations 7 and 14 and of 7 and 15 as $\sigma_{\epsilon 1 \upsilon}$ and $\sigma_{\epsilon 2 \upsilon}$ respectively. Their signs reveal whether, given that the X_i, observations are positively or negatively selected into levels of Y_2. If $\sigma_{\epsilon 2 \upsilon} > 0$, then observations are positively selected into the condition $Y_2 = 1$, and if $\sigma_{\epsilon 1 \upsilon} < 0$, then observations are positively selected into the condition $Y_2 = 0$. The covariances reveal whether the *regime* of sorting observations into classes follows the principle of, for example, comparative advantage (which holds if observations are positively selected into both groups) or some other principal.

Gamoran & Mare (1989) use endogenous switching models to examine the effects of academic tracking on the achievement of high school students when they are nonrandomly assigned to tracks. In these models Equations 14 and 15 predict the academic achievement levels of students in non-college and college bound tracks, respectively. The independent variables include the students' prior levels of achievement and social backgrounds. Each student is viewed as having two possible achievement outcomes, namely, achievement were he or she assigned to the college track Y^1_{1i}, and achievement were he or she assigned to the noncollege track Y^0_{1i}. In fact, however, each student is observed in only one track, and her or his (expected) achievement in the other track is censored. Equation 7 represents the process by which students are assigned to the college or non-college tracks. Gamoran & Mare consider two forms of Equation 7, a structural form and a reduced form. In the structural form, the independent variables include not only the social background and prior achievement levels of students but also their expected levels of achievement in the two tracks. This represents the idea that track assignment decisions, whether made by parents, school officials, or the students themselves, may be affected by expectations of how well a student will perform in alternative tracks. In practice, expected levels of achievement are only partially observed and it is necessary to solve for the reduced form of Equation 7, which includes not only the exogenous predictors of track assignment but also, subject to some constraints, the determinants of expected achievement, as represented in 14 and 15. Equations 14, 15, and the reduced form of 7 are estimated jointly. By modelling track assignment, one can take account of the selection bias that may occur if Equations 14 or 15 were estimated alone. Conversely, the model allows one also to explore how expected achievement may affect track assignment. By placing restrictions on the models, one can test alternative ideas about the ways that schools and families make tracking decisions (Gamoran & Mare 1989, Mare & Winship 1988).

In other applications of endogenous switching models, Sakamoto & Chen (1991) assess the effects of labor market sector on earnings taking into account the nonrandom allocation of workers to sectors; Willis & Rosen (1979) examine the effects of college attendance on earnings in a model for the self-selection of students with varying abilities to alternative levels of schooling; and Manski et al (1992) estimate the effects of being raised in a female-headed family on high school graduation, using a variant of the treatment model to take account of self-selection of individuals into family statuses.

ESTIMATORS

A large number of estimators have been proposed for selection models. Until recently, all of these estimators made strong assumptions about the distribution of errors. Two general classes of methods, maximum likelihood and nonlinear least squares, typically assume bivariate normality of ϵ_i and v_i. The most popular method is that of Heckman (1976a, 1979), which only assumes that v_i in equation (7) is normally distributed and $E[\epsilon_i \mid v_i]$ is linear. Computer software packages such as *LIMDEP* (Greene 1990) implement a number of estimation strategies and selection models, including cases where the selection equation is a tobit model, a multinomial logit, or multiple criteria model, and the structural equation is a linear, probit, or tobit model.

Recently researchers have been concerned with the sensitivity of the Heckman estimator to the normality and linearity assumptions. Because maximum likelihood and nonlinear least squares make even stronger assumptions, they are typically more efficient (Nelson 1984) but even less robust to violations of distributional assumptions. This lack of robustness is also a property of Olson's (1980) linear probability estimator which assumes that errors are uniformly as opposed to normally distributed. The main concern of the recent literature is the search for alternatives to the Heckman estimator that do not depend on normality and linearity assumptions. Thus we do not review the estimators that make stronger assumptions (Maddala 1983, Amemiya 1985). Instead we first describe the Heckman estimator, discuss the concerns with its sensitivity, and review alternatives that have been proposed.

Heckman's Estimator

The Heckman estimator involves (*a*) estimating the selection model (equations 7–9); (*b*) calculating the expected error, $\hat{v}_i = \mathrm{E}[v_i \mid v_i > -Z_i\alpha]$, for each observation using the estimated α; and (*c*) using the estimated error as a regressor in 1. We can rewrite Equation 10 as:

$$Y_{1i} = X_i\beta + \mathrm{E}(\epsilon_i \mid v_i > -Z_i\alpha) + \eta_i. \qquad 16.$$

If ϵ_i and v_i are bivariate normal and $\mathrm{Var}(v_i) = 1$ then $\mathrm{E}(\epsilon_i \mid v_i) = \sigma_{\epsilon v} v_i$ and

$$\mathrm{E}(\epsilon_i \mid v_i > -Z_i\alpha) = \sigma_{\epsilon v}\, \phi(-Z_i\alpha)/[1\text{-}\Phi(-Z_i\alpha)] = \sigma_{\epsilon v}\, \lambda(-Z_i\alpha) \qquad 17.$$

where ϕ and Φ are the standardized normal density and distribution functions respectively. The ratio $\lambda(-Z_i\alpha)$ is the inverse Mills' ratio. Substituting Equation 17 into 16 we get:

$$Y_{1i} = X_i\beta + \sigma_{\epsilon v}\, \lambda(-Z_i\alpha) + \eta_i \qquad 18.$$

where η_i is uncorrelated with both X_i and $\lambda(-Z_i\alpha)$. The assumption that ϵ_i and v_i follow a bivariate normal distribution is needed: (*a*) to obtain a linear relationship between ϵ_i and v_i and (*b*) to obtain a marginally normal error v_i which produces the Mills ratio formula. No other properties of the bivariate normal are used in arriving at equation 18. In particular, no assumptions are needed about the marginal distribution of ϵ_i or its higher moments. This contrasts with the method of maximum likelihood (e.g. Amemiya 1985), which makes stronger assumptions.

The steps in Heckman's estimator are: (*a*) estimate α in 7 using a probit model; (*b*) use the estimated $Z_i\hat{\alpha}$ to calculate $\hat{\lambda}(Z_i\hat{\alpha}) = E[v_i \mid v_i > -Z_i\hat{\alpha}] = \phi(-Z_i\alpha)/[1 - \Phi(-Z_i\alpha)]$; and (*c*) estimate β and $\sigma_{\epsilon v}$ in (18) by replacing $E[v_i \mid v_i > -Z_i\alpha]$ with $\lambda(-Z_i\alpha)$. Estimation of Equation 18 by OLS gives consistent parameter estimates, but special formulas are needed to get correct standard errors because the errors, η_i, are heteroskedastic and correlated (Heckman 1979, Maddala 1983).

The precision of the estimates in Equation 18 is sensitive to the variance of λ and collinearity between X and λ. The variance of λ is determined by how effectively the probit equation at the first stage predicts which observations are selected into the sample. The better the prediction, the greater the variance of λ, and the more precise estimates will be. Collinearity will be determined in part by the overlap in variables between X and Z. If X and Z are identical, then the model is only identified because λ is nonlinear. Since it is seldom possible to justify the form of λ on substantive grounds, successful use of the method usually requires that at least one variable in Z not be included in X. Even in this case X and $\lambda(-Z\alpha)$ may be highly collinear leading to imprecise estimates.

Robustness of Heckman's Estimator

Because of the sensitivity of Heckman's estimator to model specification, researchers have focussed on the robustness of the estimator to violations of its several assumptions. Estimation of 7–9 as a probit model assumes that the errors v_i are homoskedastic. When this assumption is violated, the Heckman procedure yields inconsistent estimates (Arabmazar & Schmidt 1981), though procedures are available to correct for heteroskedasticity (Hurd 1979).

The assumed bivariate normality of v_i and ϵ_i in the selection model is needed in two places. First, normality of v_i is needed for consistent estimation of α in the probit model. Second, the normality assumption implies a particular nonlinear relationship for the effect of $Z_i\alpha$ on Y_{2i} through λ. If the expectation of ϵ_i conditional on v_i is not linear and/or v_i is not normal, λ misspecifies the relationship between $Z_i\alpha$ and Y_{2i} and the model may yield biased results.

Several studies have analytically investigated the bias in the single equation (tobit) model when the error is not normally distributed. In a model with only an intercept–that is, a model for the mean of a censored distribution—when errors are not normally distributed, the normality assumption leads to substantial bias. This result holds even when the true distribution is close to the normal (for example, the logistic) (Goldberger 1983). When the normality assumption is wrong, moreover, maximum likelihood estimates may be worse than simply using the observed sample mean. For samples that are 75% complete, bias from the normality assumption is minimal; in samples that are 50% complete, the bias is substantial in the truncated case, but not the censored; and in samples that are less than 50% complete, it is substantial in almost all cases (Arabmazar & Schmidt 1982).

That estimation of the mean is sensitive to distributional misspecification suggests that the Heckman estimator may not be robust and raises the question of how commonly such problems arise in practice. In addition, even when normality holds, the Heckman estimator may not improve the mean square error of OLS estimates of slope coefficients in small samples (50) (Stolzenberg & Relles 1990). This appears to parallel the standard result that when the effect of a variable is measured imprecisely, inclusion of the variable may enlarge the mean square error of the other parameters in the model (Leamer 1983).

No empirical work that we know of directly examines the sensitivity of Heckman's method for a standard selection model. However, several recent studies (Lalonde 1986, Lalonde & Maynard 1987) evaluate the closely related methods for assessing the impact treatment effects in nonexperimental settings. The models are made up of two equations, one predicting whether an individual participates in, for example, a job training program, and the other providing the effects of the program and other regressors on the individual's wages. This work compares the estimates of the effects of programs on wages from applying OLS and variants of Heckman's methods to nonexperimental data to estimates from data where individuals are randomly assigned to treatment conditions. Compared to OLS, Heckman's estimates yield program effects that are closer to the experimental results. The Heckman estimates, however, often differ substantially from the experimental estimates and tend to fluctuate depending on which variables are included in the selection equation. The various estimates reported by Lalonde have large standard errors, mainly because of small samples; hence they are not a definitive appraisal of the Heckman methods. But this research strongly suggests that Heckman's method is no panacea for selection problems and, when its assumption are not met, may yield misleading results.

Extensions of the Heckman Estimator

There are two main issues in estimating Equation 18. First, is the equation that predicts selection into the sample consistently estimated? That is, are estimates of α, which derive from the selection equation, consistent? This depends on the assumptions that (*a*) the independent variables in that equation (Z_i) have linear effects, and (*b*) the errors in the selection equation are normally distributed. Assumption (*a*) depends on the same considerations as any linear model as to whether interactions or other nonlinear transformations of the regressors should be included. Unfortunately, a strong substantive rationale for the regressors included in the selection equation is often unavailable. Likewise, assumption (*b*) in practice seldom rests on a firm substantive basis.

Second, what nonlinear function should be chosen for λ, which dictates how the predicted probabilities of sample selection affect the dependent variable in Equation 18? When bivariate normality of errors holds, λ is the inverse Mills ratio. When this assumption does not hold, inconsistent estimates may result. Moreover, since the regressors in the main and sample selection equations (X_i and Z_i) are often highly collinear, estimates of β in (18) may be sensitive to misspecification of λ.

Much of the recent research on models for selection has focussed on developing estimators that do not rely on these distributional and functional form assumptions. Most work thus far is theoretical, although some applications of these new methods have been carried out. Many of the new estimators relax the assumptions of Heckman's two-step approach. The approach of many of the new models is as follows. First, the selection model, Equations 7–9, is estimated using a nonparametric method for binary regression models. These methods include Manski's maximum score method (1975, 1985), which is implemented in *LIMDEP,* nonparametric maximum likelihood estimation (Cosslett 1983), weighted average derivatives (Stoker 1986, Powell et al 1989), and kernel estimation (Bierens 1990, Ichimura 1988, Klein & Spady 1987). Spline methods and series approximations (Hardle 1990) are also available but are, as far as we are aware, an unexplored approach. Two bases for evaluating these methods are (*a*) the trade-off that they make between efficiency and the strength of their prior assumptions, and (*b*) their asymptotic distribution. Chamberlain (1986) establishes a theoretical upper bound for the efficiency of nonparametric methods under particular assumptions. Kernel methods (Ichimura 1988, Klein & Spady 1987) and variants of the weighted average derivatives (Stoker 1991) reach this bound, but other methods that make weaker assumptions, such as the method of scoring (Manski 1975), do not. For still others, the efficiency is unknown (Cosslett 1983). We discuss some of the assumptions made in alternative semiparametric and nonparametric approaches below. Asymptotic normality has been established for all these estimators except those of Manski and Cosslett.

Kernel Estimation

As far as we are aware, kernel estimation is the only nonparametric approach that has been used for the first stage of selection models in empirical applications. This approach is as follows. Asssume that we have multiple observations of Y_2 for each possible value of the vector Z_i. Let $g(Z)$ be the function for the conditional mean of Y_2 on Z. Then a nonparametric estimator of the conditional mean functions is:

$$\hat{g}(Z) = \Sigma_i Y_{2i} \qquad \text{for all } Z_i = Z.$$

For example, if we were predicting whether individuals were employed or not (Y_{2i}) from their level of educational attainment (Z_i), our nonparametric estimate would simply be the proportion of persons who are employed at each level of educational attainment. This estimator makes no assumption about how Z enters g, (for example, that it is linear, making the model semiparametric), or about the distribution for the error v (which would make the model parametric). If data are grouped so that multiple values of Y_{2i} are known for each value of Z_i, this procedure is straightforward. If, however, Z_i varies continuously, so that there is at most one observation of Y_{2i} at each value of Z_i, kernel estimation is required.

The kernel method uses observations (Y_{2j}, Z_j) where Z_j is close to Z_i to estimate the mean of Y_{2i}. We assume that g_i is continuous and calculate a weighted average of the Y_{2i} to estimate g_i, where observations with Z_j that are close to Z_i are weighted more heavily than observations that are further away; that is, $\hat{g}_i = \Sigma_j K_{i\,j} Y_{2j} / \Sigma_j K_{ij}$, where $K_{ij} = K[(Z_i - Z_j)/h]$. K is assumed to have a maximum at zero and to decrease as the absolute size of $Z_i - Z_j$ increases. Although many functions are possible for K, the choice of function does not usually affect the estimates. The researcher selects h, known as the bandwidth of K, to control the smoothness of the estimator (Hardle 1990). As the sample increases, h should gradually approach zero, which guarantees a consistent estimate of g_i.

As in Heckman's method, the second stage is to estimate Equation 18, using the estimates of α or g_i from the first stage. Several approaches are available to estimate 18 without parametrically specifying λ. One approach is to approximate λ through a series expansion (Newey 1990) such as Edgeworth series (Lee 1982), or by step functions (Cosslett 1991), with the number of terms gradually increasing with sample size. A second possibility is to use kernel methods to estimate λ (Robinson 1988, Powell 1987, Ahn & Powell 1990). By one interpretation, this is a generalized difference estimator (Powell 1987). In this approach one differences out the nuisance function $\lambda(g_i)$ by estimating Equation 18 across differences between pairs of observations:

$$Y_{1i} - Y_{1j} = (X_i - X_j)\beta + [\lambda(g_i) - \lambda(g_j)] + \epsilon_i - \epsilon_j. \qquad 19.$$

If one only uses pairs for which the probability of selection is equal ($g_i = g_j$), then the terms in λ simply drop out of Equation 19 and OLS can be used. If λ

is continuous, for pairs i, j for which $g_i \cong g_j$, $\lambda(g_i) \cong \lambda(g_j)$, and the $[\lambda(g_i) - \lambda(g_j)]$ will be near zero. Powell's procedure uses all pairs and weights more heavily pairs for which the difference $g_i - g_j$ is less. As the sample increases, more weight is placed on pairs for which $g_i \cong g_j$, thus guaranteeing consistency of the estimator. Powell's approach will only identify the effects of Xs that vary across individuals with the similar g. As a result, it is not possible to identify the intercept using his approach. Estimates of intercepts may be important in both the treatment and endogenous switching models.

Empirical Applications

All three of these methods have been compared to Heckman's normal estimator using a common model on a single set of data. Several investigators have used Mroz's (1987) model on the labor force participation of married women. In this model, the dependent variable is annual hours worked; a selection equation models whether or not a woman worked more than zero hours. Mroz uses Heckman's two-stage method for estimating the model. Newey et al (1990) and Ahn & Powell (1990) use semi- and nonparametric methods. These studies provide a common context for comparing (*a*) Heckman's method, (*b*) weighted kernel estimation of the hours equation and probit estimates of the selection equation, (*c*) series expansion estimation of λ with probit estimates of the selection equation, and (*d*) weighted kernel estimation of both the hours and the selection equations. Methods (*b*) and (*c*) are semiparametric; method (*d*) is nonparametric.

The weighted kernel and series expansion results are generally similar to those from Heckman's method although their standard errors are typically slightly larger. When kernel methods are used for both the hours and selection equations, some coefficients differ markedly from the other methods, and the standard errors are much larger than for the other methods. Nonparametric estimates for the selection equation are very imprecise. At least in this example, moreover, results are sensitive to alternative estimation approaches. The nonparametric procedures make the weakest assumptions, but their standard errors are so large as to imply that the data are consistent with a very wide range of results.

Manski's Bound Approach

Although semi- and nonparametric methods are conservative, they are not free of assumptions. For example, in Ahn & Powell's model, λ is assumed to be a function of a single index, g_i, and to enter Equation 18 additively. Without such assumptions it is often impossible even to put a bound on the conditional mean of Y_1 given X, the usual quantity estimated in regression analysis, much less obtain a consistent point estimate. In an important set of papers Manski (1989, 1990, 1991) shows that, without prior assumptions about the selection process, it is impossible to obtain point estimates of the true regression model when selection occurs. Under some conditions, however, it is possible to place bounds on the estimated regression coefficients in the absence of assumptions about the selection process. A regression model can be written as:

$$E(Y_1 \mid X) = E(Y_1 \mid X, Y_2 = 1)P(Y_2 = 1) + E(Y_1 \mid X, Y_2 = 0)P(Y_2 = 0), \quad 20.$$

where all of the notation is as defined above. All of the components in Equation 20 can be estimated from observed data except $E(Y_1 \mid X, Y_2 = 0)$, the regression of Y_1 on X for cases that are not in the selected sample. Unless one can put bounds on this value, sample data provide no information about the true regression of Y_1 on X. Manski derives bounds for $E(Y_1 \mid X, Y_2 = 0)$ when Y_1 is dichotomous and for conditional *medians* when Y_1 is continuous (Manski 1989, 1990, 1991). In general the tightness of the bound varies inversely with the proportion of cases in the sample that are censored. Somewhat surprisingly, Manski shows that it is easier to obtain bounds for regression estimates when the dependent variables are dichotomies or medians than when they are continuous. Manski has written a computer program for doing this type of analysis. Although researchers are in the habit of seeking a point estimate for parameters of interest, this usually comes at the price of making questionable identifying assumptions. Manski shows that these assumptions can be relaxed if one is willing to settle for an informative range of possible parameter values.

Methods Based on Panel Data

Panel data are useful for estimating treatment effects when subjects are nonrandomly assigned to conditions, and also enable one to take account of some kinds of nonrandom sample selection (Heckman & Robb 1985, 1986a,b; Heckman & Hotz 1989). If selection is only on the observed independent variables, then selection bias is not a problem. Panel data enable one to control some unobserved as well as observed variables and may in some cases alleviate selection bias. We consider models with fixed, observation-specific effects, although the approach can be generalized to models with random effects and, in some cases, with time-varying effects. The model is:

$$Y_{1it} = X_{it}\beta + v_i + \epsilon_{it} \quad 21.$$

where t indexes time and v_i is an unobserved component that is unique to each cross-sectional observation but is invariant over time. The selection equation is:

$$Y_{2it} = Z_{it}\alpha + \gamma v_i + \eta_{it}, \tag{22}$$

where the selection rule is given by Equations 5 and 6 above. If ϵ_{it} and η_{it} are uncorrelated, and we can take account of v_i, then selection is on the observed independent variables alone and Equation 21 is estimable by OLS. A method of taking account of v_i is to take deviations of the variables in 21 from their observation-specific means. One can estimate:

$$Y_{1it} - \bar{Y}_{1i} = (X_{1it} - \bar{X}_{1i})\beta + (\epsilon_{it} - \epsilon_i), \tag{23}$$

which yields estimates of β in 21 because v_i is invariant over time. This eliminates v_i and thus the selection problem. The key assumption is that the unobserved determinants of selection are time invariant characteristics of the observations. If the unobserved determinants of selection depend on time, then a more complex model is required.

CONCLUSION

Infallible models for sample selection bias do not exist. Methods are rapidly evolving and, at present, different methods may yield different results. We do not know definitively the robustness of Heckman's estimator and its generalizations across different empirical contexts. We do know that, in some contexts, methods that make weaker assumptions give different results from Heckman's estimator, but the weaker assumptions may come at the cost of greatly reduced precision. When selection is an issue, therefore, empirical results are likely to remain ambiguous (Manski 1989). What should the researcher do? Because one's results may depend on the method used, researchers should be explicit about the assumptions behind their methods and should present estimates using a variety of methods. Manski et al (1992) exemplify this approach in their analysis of the effects of family structure on the likelihood dropping out of high school, which includes parametric and nonparametric models, and an analysis of bounds.

Acknowledgment

The research was supported by grants from the National Institute of Child Health and Human Development and the National Science Foundation. The authors are grateful to Charles Manski, Peter Marsden, John Hagan, and Tom Cook for helpful discussions and comments on an earlier draft. The authors also appreciate the bibliographic assistance of Ruth Sandor and Carol Sleeva and the assistance of Barbara Corry and Nancy Bennett in preparing the manuscript.

Literature Cited

Ahn, H., Powell, J. J. 1990. Semiparametric estimation of censored selection models with a nonparametric selection mechanism. Madison, Wisc: Dep. Econ., Univ. Wisc.

Amemiya, T. 1985. *Advanced Econometrics*. Cambridge, Mass: Harvard Univ. Press

Arabmazar, A., Schmidt, P. 1981. Further evidence of robustness of the tobit estimator to heteroscedasticity. *J. Economet.* 17:253–58

Arabmazar, A., Schmidt, P. 1982. An investigation of the robustness of the Tobit estimator to non-normality. *Econometrica* 50:1055–63

Ashenfelter, 0. 1978. Estimating the effect of programs on earnings. *Rev. Econ. Stat.* 60:47–57

Barnett, W. A., Powell, J., Tauchen, G. 1991. *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge: Cambridge Univ. Press

Barnow, B. S., Cain, G. G., Goldberger, A. S. 1980. Issues in the analysis of selectivity bias. In *Eval. Stud. Rev.*, 5:43–59, ed. E. W. Stromsdorfer, G. Farkas. Beverly Hills, Calif: Sage

Berk, R. A. 1983. An introduction to sample selection bias in sociological data. *Am. Sociol. Rev.* 48:386–98

Berk, R. A. 1988. Causal inference for sociological data. In *The Handbook of Sociology,*, ed. N. J. Smelser, pp. 155–72. Newbury Park, Calif: Sage

Berk, R. A., Ray, S. C. 1982. Selection biases in sociological data. *Soc. Sci. Res.* 11:352–98

Bierens, H. 1990. Kernel estimators of regression functions. In *Advances in Econometrics 1990,* ed. T. Bewley, pp. 99–144. Cambridge, UK: Cambridge Univ. Press

Chamberlain, G. 1986. Asymptotic efficiency in semi-parametric models with censoring. *J. Economet.* 32:189–218

Cosslett, S. R. 1983. Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica* 51:765–81

Cosslett, S. R. 1991. Semiparametric estimation of a regression model with sampling selectivity. In *Nonparametric and Semiparametric Methods in Econometrics and Statistics,*, ed. W. A. Barnett, J. Powell, G. Tauchen, pp. 175–98. Cambridge: Cam bridge Univ. Press

DuMouchel, W. H., Duncan, G. J. 1983. Using sample survey weights in multiple regression analyses of stratified samples. *J. Am. Stat. Assoc.* 78:535–43

England, P., Farkas, G., Kilbourne, B., Dou, T. 1988. Explaining occupational sex segregation and wages: findings from a model with fixed effects. *Am. Sociol. Rev.* 53:544–58

Fligstein, N. D., Wolf, W. 1978. Sex similarities in occupational status attainment: are the results due to the restriction of the sample to employed women? *Soc. Sci. Res.* 7:197–212

Gamoran, A., Mare, R. D. 1989. Secondary school tracking and stratification: compensation, reinforcement, or neutrality? *Am. J. Sociol.* 94:1146–83

Goldberger, A. S. 1981. Linear regression after selection.*J. Economet.* 15:357–66

Goldberger, A. S. 1983. Abnormal selection bias. In *Studies in Econometrics Time Series, and Multivariate Statistics*, ed. S. Karlin, T. Amemiya, L. A. Goodman, pp. 67–84. New York: Academic Press

Greene, W. H. 1990. *LIMDEP*. New York: Econometric Software

Gronau, R. 1974. Wage comparisons—selectivity bias. *J. Polit. Econ.* 82:1119–43

Grusky, D. B. 1986. *American Social Mobility in the 19th and 20th Centuries. CDE Working Paper 86–28*. Madison: Cent. Demogr. Ecol., Univ. Wisc.-Madison

Hagan, J. 1990. The gender stratification of income inequality among lawyers. Soc Forces 68:835–55

Hagan, J., Parker, P. 1985. White-collar crime and punishment. *Am. Sociol. Rev.* 50:302–16

Hannan, M. T., Tuma, N. B., Groeneveld, L. P. 1977. Income and marital events: evidence from an income-maintenance experiment. *Am. J. Sociol.* 82:1186–211

Hannan, M. T., Tuma, N. B., Groeneveld, L. P. 1978. Income and independence effects on marital dissolution: results from the Seattle and Denver income- maintenance experiments. *Am. J. Sociol.* 84:611–34

Hardle, W. 1990. *Applied Nonparametric Regression*. New York: Cambridge Univ. Press

Hardy, M. A. 1989. Estimating selection effects in occupational mobility in a 19th-century city. *Am. Sociol. Rev.* 54:834–43

Hausman, J. A., Wise, D. A. 1976. The evaluation of results from truncated samples: The New Jersey income maintenance experiment, *Ann. Econ. Soc. Measurement* 5/4:421–45

Hausman, J. A., Wise, D. A. 1977. Social experimentation, truncated distributions, and efficient estimation. *Econometrica* 45: 919–38

Heckman, J. J. 1974. Shadow prices, market wages and labor supply. *Econometrica* 42: 679–94

Heckman, J. J. 1976a. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Ann. Econ. Soc. Measurement*. 5:475–92

Heckman, J. J. 1976b. Simultaneous equation models with continuous and discrete endogenous variables and structural shifts. In *Studies in Nonlinear Estimation,* ed. S. Goldfeld, R. Quandt, pp. 235–72. Cambridge, Mass: Ballinger

Heckman, J. J. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46:931–59

Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica* 47:153–61

Heckman, J. J. 1987. Selection bias and self-selection. In *The New Palgrave,* ed. J. Eatwell, M. Milgate, P. Newmann, pp. 287–97. New York: Stockton

Heckman, J. J., Robb, R. 1985. Alternative methods for evaluating the impact of interventions. In *Longitudinal Analysis of Labor Market Data,* ed. J. J. Heckman, B. Singer, pp. 156–245. Cambridge: Cambridge Univ. Press

Heckman, J. J., Sedlacek, G. 1985. Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market. *J. Polit. Econ.* 93:1077–25

Heckman, J. J., Robb, R. 1986a. Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In *Drawing Inferences from Self-Selected Samples,* ed. H. Wainer, pp. 63–107. New York: Springer- Verlag

Heckman, J. J., Robb, R. 1986b. Alternative identifying assumptions in econometric models of selection bias. Adv. Econ. 5:243–87

Heckman, J. J., Hotz, V. J. 1989. Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training. *J. Am. Stat. Assoc.* 84:862–74

Heckman, J. J., Honore, B. E. 1990. The empirical content of the Roy model. *Econometrica* 58:1121–50

Heckman, J. J., Sedlacek, G. 1990. Self-selection and the distribution of hourly wages. *J. Labor Econ.* 8:S329–63

Hoffman, P. T. 1984. Wills and statistics: tobit analysis and the counter reformation in Lyon. J. *Interdis. Hist.* 14:813–34

Holland, P. W. 1986. Statistics and causal inference. *J. Am. Stat. Assoc.* 81:945–70

Hurd, M. 1979. Estimation in truncated samples. *J. Economet.* 11:247–58

Ichimura, H. 1988. *Semiparametric Least Squares Estimation of Single Index Models. Technical Report*. Minneapolis: Dept. Econ., Univ. Minn., Minneapolis

Kalbfleisch, J. D., Prentice, R. L. 1980. *The Statistical Analysis of Failure Time Data.* New York: Wiley

Klein, R. W., Spady, R. S. 1987. An efficient semiparametric estimator of the binary response model. Morristown, NJ: Bell Commun. Res.

Lalonde, R. J. 1986. Evaluating the econometric evaluations of training programs with experimental data. *Am. Econ. Rev.* 76:604–20

Lalonde, R., Maynard, R. 1987. How precise are evaluations of employment and training programs: evidence from a field experiment. *Eval. Rev.* 11:428–51

Lancaster, T. 1990. *The Econometric Analysis of Transition Data*. Cambridge: Cambridge Univ. Press

Learmer, E. E. 1983. Model choice and specification analysis. In *Handbook of Econometrics,* Vol. 1, ed. Z. Griliches, M. D. Intriligator, pp. 284–327. Amsterdam: North-Holland

Lee, L. F. 1982. Some approaches to the correction of selectivity bias. *Rev. Econ. Stud.* 49:355–72

Lee, L. F. 1983. Generalized econometric models with selectivity. *Econometrica* 51: 507–12

Lewis, H. G. 1974. Comments on selectivity biases in wage comparisons. *J. Polit. Econ.* 82:1145–56

Lieberson, S. 1985. *Making it Count: The Improvement of Social Research and Theory*. Los Angeles: Univ. Calif. Press

Little, R. J. A., Rubin, D. B. 1987. *Statistical Analysis with Missing Data*. New York: Wiley

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

Manski, C. F. 1975. Maximum score estimation of the stochastic utility model of choice. *J. Economet.* 3:205–28

Manksi, C. F. 1985. Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator. *J. Economet.* 27:313–33

Manski, C. F. 1989. Anatomy of the selection problem. *J. Hum. Resources* 24:343–60

Manski, C. F. 1990. Nonparametric bounds on treatment effects. *Am. Econ. Rev.* 80: 319–23

Manski, C. F. 1991. The selection problem. In *Advances in Econometrics 1990,* ed. C. Simms. New York: Cambridge Univ. Press

Manski, C. F., Wise, D. A. 1983. *College Choice in America*. Cambridge, Mass: Harvard Univ. Press

Manski, C. F., Garfinkel, I., eds. 1992. *Evaluating Welfare and Training Pro-*

grams. Cambridge, Mass: Harvard Univ. Press

Manski, C. F., McLanahan, S. S., Powers, D., Sandefur, G. D. 1992. Alternative estimates of the effects of family structure during adolescence on high school graduation. *J. Am. Stat. Assoc*. 54:25–37

Mare, R. D., Winship, C. 1984. The paradox of lessening racial inequality and joblessness among black youth: enrollment, enlistment, and employment, 1964–1981. *Am. Sociol. Rev*. 49:39–55

Mare, R. D., Winship, C., Kubitschek, W. N. 1984. The transition from youth to adult: understanding the age pattern of employment. *Am. J. Sociol.*. 89:326–58

Mare, R. D., Chen, M. D. 1986. Further evidence on number of siblings and educational stratification. *Am. Sociol. Rev*. 51:403–12

Mare, R. D., Winship, C. 1988. Endogenous switching regression models for the causes and effects of discrete variables. In *Common Problems/Proper Solutions: Avoiding Error in Quantitative Research,* ed. J. S. Long, pp. 132–60. Newbury Park, Ca lif: Sage

Marini, M. M., Singer, B. 1988. Causality in the social sciences. In *Sociological Methodology,* ed. C. C. Clogg, pp. 347–411. San Francisco: Jossey-Bass

Marsden, P. V., Hurlbert, J. S. 1987. Small networks and selectivity bias in the analysis of survey network data. *Soc. Networks* 9:333–49

Mroz, T. A. 1987. The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica* 55:765–99

Nelson, F. D. 1984. Efficiency of the two-step estimator for models with endogenous sample selection. *J. Economet*. 24:181–96

Newey, W. K. 1990. *Two-step series estimation of sample selection models*. Pap. presented 1988 European Meet. Econometrica Soc.

Newey, W. K., Powell, J. L., Walker, J. R. 1990. Semiparametric estimation of selection models: some empirical results. *Am. Econ. Rev*. 80:324–28

Olson, R. J. 1980. A least squares correction for selectivity bias. *Econometrica* 48:1815–20

Peterson, R., Hagan, J. 1984. Changing conceptions of race: towards an account of anomalous findings of sentencing research. *Am. Sociol. Rev*. 49:56–70

Powell, J. L. 1987. *Semiparametric Estimation of Bivariate Latent Variable Models. Working Paper No. 8704*. Madison, Wisc: Soc. Systems Res. Inst., Univ. Wisc.

Powell, J. L., Stock, J. H., Stoker, T. M. 1989. Semiparametric estimation of index coefficients. *Econometrica* 57:1403–30

Robinson, P. M. 1988. Root-N-Consistent semiparametric regression. *Econometrica* 56:931–54

Rosenbaum, P., Rubin, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55

Roy, A. D. 1951. Some thoughts on the distribution of earnings. *Oxford Econ. Pap*. 3:135–46

Rubin, D. B. 1978. Bayesian inference for causal effects: the role of randomization. *Ann. Statist*. 6:34–58

Ruud, P. 1986. Consistent estimation of limited dependent variable models despite misspecifications of distribution. *J. Economet*. 32:157–87

Sakamoto, A., Chen, M. D. 1991. Inequality and attainment in the dual labor market *Am. Sociol. Rev*. 56:295–308

Seltzer, J. A. 1991. Legal custody arrangements and children's economic welfare. *Am. J. Sociol*. 96:895–929

Seltzer, J. A., Garfinkel, I. 1990. Inequality in divorce settlements: an investigation of property settlements and child support awards. *Soc. Sci. Res*. 19:82–111

Steelman, L. C., Powell, B. 1989. Acquiring capital for college: the constraints of family configuration. *Am. Sociol. Rev*. 54:844–55

Stoker, T. M. 1986. Consistent estimates of scaled coefficients. *Econometrica* 54:1461-81

Stoker, T. M. 1991. Equivalence of direct, indirect, and slope estimators of average derivatives. In *Nonparametric and Semiparametric Methods in Econometricas and Statistics,* ed. W. A. Barnett, J. Powell, G. Tauchen, pp. 99–118. Cambridge: Cambridge Univ. Press

Stolzenberg, R. M., Relles, D. A. 1990. Theory testing in a world of constrained research design: the significance of Heckman's censored sampling bias correction for nonexperimental research. *Sociol. Meth. Res*. 18:395–415

Tobin, J. 1958. Estimation of relationships for limited dependent variables. *Econometrica* 26:24–36

Wainer, H., ed. 1986. *Drawing Inferences from Self-Selected Samples*. New York: Springer-Verlag

Willis, R- J., Rosen, S. 1979. Education and self-selection. *J. Polit. Econ*. 87:S507–36

Zatz, M. S., Hagan, J. 1985. Crime, time, and punishment: an exploration of selection bias in sentencing research. *J. Quant. Criminol*. 1:103–26