working paper
department
of economics

CONDITIONAL INDEPENDENCE IN SAMPLE SELECTION MODELS

Joshua D. Angrist

96-27 **October, 1996**

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

CONDITIONAL INDEPENDENCE IN SAMPLE SELECTION MODELS

Joshua D. Angrist

96-27 **October, 1996**

Conditional Independence in Sample Selection Models

by,

Joshua D. Angrist

MIT, Hebrew University, and NBER*

October 1996

*Thanks go to Guido Imbens, Whitney Newey, Paul Rosenbaum, Tom Stoker, and especially to Gary Chamberlain for helpful discussions and comments. Any errors are, of course, solely the work of the author.

Conditional Independence in Sample Selection Models

Econometric sample selection models typically use a linear latent-index with constant coefficients to model the selection process and the conditional mean of the regression error in the selected sample. A feature common to most of these models is that the conditional mean function for regression errors is an invertible function of the selection propensity score, i.e., the probability of selection conditional on covariates. Consequently, conditioning on the selection propensity score controls selection bias, a fact which underlies much of the recent literature on non-parametric and semi-parametric selection models. This literature has not addressed the question of whether the propensity-score conditioning property is necessarily a feature of sample selection models. In this note, I describe the conditional independence properties that make it possible to use the selection propensity score to control selection bias in a general sample selection model. The resulting characterization does not rely on a latent index selection mechanism and imposes no structure other than an assumption of independence between the regression error term and the regressors in random samples. This approach leads to a simple rule that can be used to determine if conditioning on the selection propensity score is sufficient to control selection bias.

Joshua D. Angrist
MIT Department of Economics
50 Memorial Drive
Cambridge, MA 02139-4307

ANGRIST@MIT.EDU

Keywords: non-random samples, selection bias, propensity score

1. Introduction

Sample selection problems occur in both experimental and observational studies. Econometricians' interest in sample selection models originated with research in labor economics using the distribution of wages for workers to infer the distribution of offered wages for everyone (Gronau, 1974; Heckman, 1974). A similar problem occurs when instrumental variables that are being used to correct for endogeneity in wage equations also affect employment status (see, e.g., Angrist, 1995). In experiments, the bias induced by controlling for "post-treatment variables" that are correlated with the outcome of interest can also be interpreted as a form of selection bias (Ham and Lalonde, 1996; Rosenbaum, 1984).

The traditional econometric solution to problems of this type invokes an assumption of joint normality for the regression error and the error term in a latent-index selection mechanism. This approach leads to a maximum likelihood or two-step selection correction in the form of inverse Mills-ratio terms added to the regression function of interest (see Heckman, 1976 and 1979). A number of less restrictive semi-parametric variations on this normal selection model have also been introduced (see, e.g., Ahn and Powell, 1993, and references in Newey, Powell, and Walker, 1990 and Heckman, 1990). Both the parametric and semi-parametric formulations have two common threads. First, a latent index framework is used to characterize the mean of unobserved regression error terms conditional on the regression covariates and the sample selection rule. Second, the problem of selection bias is either implicitly or explicitly handled by conditioning on an invertible function of the selection propensity score, i.e., the probability of selection given covariates.[1]

This paper describes the conditional independence properties that make it possible to use the selection propensity score to control selection bias in regression estimates. The problem studied here includes applications involving experimental data and instrumental variables as special cases. My first objective is a nonparametric formulation of the sample selection problem where the only structure consists

---

[1]The propensity-score terminology originates with Rosenbaum and Rubin (1983, 1984), who introduced the propensity-score method of controlling for confounding in evaluation research.

of the regression equation of interest and an independence restriction on the regression error. The elementary properties of conditional independence, as discussed by Dawid (1979) and Florens and Mouchart (1982), are then applied to this formulation. I believe this approach provides a very general description of the identification problem in sample selection models. It also leads to an easily interpreted rule that can be used to check when nonparametric conditioning on the selection propensity score is indeed sufficient to control selection bias.

2. Motivation

The equation of interest is

$$y_i = X_i'\beta + \varepsilon_i, \tag{1}$$

where y_i is the dependent variable and X_i is the k x 1 vector of regressors, assumed to be independent of ε_i in the population. We observe a random sample of n observations on X_i, but y_i is not observed for everybody. The coefficient vector is assumed to be such that $X_i'\beta$ gives the population conditional expectation function, so that if we did observe y_i for everyone in the sample, β could be consistently estimated by ordinary least squares (OLS). Although independence of regressors and errors is not necessary to justify this definition of β, the full independence assumption is customary in papers on selection bias (e.g., Chamberlain, 1986). At the end of the paper I show that full independence appears to be of fundamental importance for identification strategies that use the selection propensity score to control selection bias.

Let w_i indicate selection status (i.e., w_i=1 indicates workers in an application where y_i is the log of offered wages). The selection problem is often motivated using a latent index to capture the possibility that X_i affects w_i as well as y_i. In particular, suppose that

$$w_i = 1[X_i'\delta - \eta_i > 0] \tag{2}$$

where η_i is an error term that could be correlated with ε_i but is assumed to be independent of X_i. Then, we have

$$E[\varepsilon_i | X_i, w_i=1] = E[\varepsilon_i | X_i, X_i'\delta > \eta_i] \neq 0.$$

Selection bias arises because the term $E[\varepsilon_i | X_i, w_i=1]$ is generally a function of X_i even though $E[\varepsilon_i | X_i] = 0$ in random samples.

Parametric solutions to selection problems of this type typically begin by assuming that the error terms (ε_i, η_i) are jointly normally distributed, so that

$$E[\varepsilon_i | X_i, w_i=1] = \rho E[v_i | X_i, X_i'\delta > \eta_i], \tag{3}$$

where ρ is the coefficient from a regression of ε_i on v_i. Because of normality and joint independence of (ε_i, η_i) with X_i, we can simplify further:

$$E[\eta_i | X_i'\delta > \eta_i] = -\phi(X_i'\delta)/\Phi(X_i'\delta) \equiv \lambda(X_i'\delta),$$

where ϕ and Φ are the normal density and distribution functions for η_i. Adding $\lambda(X_i'\delta)$ as a regressor to (1) provides a feasible solution to the selection problem in this context because the required parameters can be estimated from a Probit regression of w_i on X_i.

Critics of traditional econometric selection models (e.g., Hartigan and Tukey, 1986; Little, 1985) point out that these models combine a wide variety of assumptions that are hard to assess and interpret. In particular, the validity of parametric corrections for sample selection bias clearly turns to a large extent on distributional assumptions, functional form restrictions, and exclusion restrictions (i.e., zero restrictions on β and/or δ) for which there is rarely, if ever, any empirical justification.[2] These features of the normal selection model have led econometricians to develop less restrictive models for selection problems. For example, Newey, Powell, and Walker (1990), Lee (1982), and others have shown that the distributional assumptions in these model can be relaxed in a semi-parametric framework.

One approach to semi-parametric selection corrections begins with the observation that even in

---

[2]Mroz (1987) documents the fact that empirical results can be sensitive to these assumptions. See also Olsen (1980, 1982).

the absence of parametric distributional assumptions, selection-correction terms typically consist of non-linear but invertible functions of the selection propensity score, $P[w_i=1 \mid Z_i]$. This feature of selection models has been noted by Heckman and Robb (1986), and it motivates the semi-parametric estimators for selection models discussed by Newey (1988), Robinson (1988), Choi (1992), and Ahn and Powell (1993).[3] On the other hand, earlier work on semi-parametric selection models has not clarified whether the propensity-score conditioning property is a <u>necessary</u> feature of the selection problem. In the next section, I use simple conditional independence notation to provide a general characterization of models for which this propensity-score conditioning property holds.

3. Conditional independence and sample selection

Conditional independence is defined as follows:

Definition 1 (Florens and Mouchart, 1982). Let A_1, A_2, A_3 denote random variables defined on a common probability space with joint probability measure $P[A_1, A_2, A_3]$, and let $P[A_1 \mid A_2, A_3]$ denote a conditional probability statement for A_1 given some values of A_2 and A_3. Then we write $A_1 \perp\!\!\!\perp A_2 \mid A_3$ if and only if $P[A_1 \mid A_2, A_3] = P[A_1 \mid A_3]$, a.s.

Note that conditional independence, like independence, is transitive. The conditional independence concept, although basic and perhaps even obvious, has proven widely useful. Dawid (1979) popularized the notation, $A_1 \perp\!\!\!\perp A_2 \mid A_3$, and argued that many important ideas in statistics can best be understood in these terms. Chamberlain (1982) and Florens and Mouchart (1982) used conditional independence ideas to explore the relationship between alternate definitions of "causality" in time-series econometrics. And

[3]Educational researchers and statisticians have also discussed conditioning on the selection propensity score; see, for example, Powell and Steelman (1984) and Wainer (1986).

Rosenbaum and Rubin (1983) used conditional independence arguments to establish the propensity-score result for evaluation problems with selection on observables.[4]

The properties of conditional independence can be characterized as follows:

Lemma 1. Let R_1, R_2, R_3, and R_4 be random variables defined on a common probability space with joint probability measure. Then the following are equivalent:

(i) $\quad R_1 \amalg R_2 \mid R_3 \quad \text{and} \quad R_1 \amalg R_4 \mid (R_2, R_3)$

(ii) $\quad R_1 \amalg (R_2, R_4) \mid R_3$

(iii) $\quad R_1 \amalg R_4 \mid R_3 \quad \text{and} \quad R_1 \amalg R_2 \mid (R_3, R_4)$

This is Theorem A.1 in Florens and Mouchart (1982), presented here in somewhat simpler notation, and Lemma 4.3 in Dawid (1979). Dawid states the equivalence of (ii) and (iii) only, but this also implies the equivalence of (i) and (iii) by using symmetry of R_2 and R_4 in part (ii).

3.1 Selection bias

The problem of selection bias can be described in conditional independence terms with no structure other than the regression of interest and the independence restriction on ε_i. This restriction is stated formally below:

Assumption 1. Let $\{(y_i, X_i')'; i=1, \ldots, n\}$ denote i.i.d. observations on a (k+1)x1 random vector. Then $\varepsilon_i \amalg X_i$ where $\varepsilon_i \equiv (y_i - X_i'\beta)$ and $E[y_i \mid X_i] = X_i'\beta$.

[4]The propensity-score result for evaluation research says that in an experiment where treatment is randomly assigned conditional on covariates, it is enough to condition on the probability of selection given covariates to eliminate confounding from any relationship between the covariates and the treatment assignment.

We would first like to know when a sample selection process causes Assumption 1 to fail. Let w_i indicate selection status as before. Instead of observing $\{(y_i, X_i')'; i=1, \ldots, n\}$, we observe $\{(w_i y_i, w_i, X_i')'; i=1, \ldots, n\}$. Important consequences of the selection process are summarized below:

Proposition 1. Suppose that X_i has at least 2 points of support and that $P[w_i=1 \mid X_i]$ is a non-trivial function of X_i. Then given Assumption 1, $\varepsilon_i \amalg X_i \mid w_i$ iff $\varepsilon_i \amalg w_i \mid X_i$.

Proof. Let Ω denote the sample space. Setting $R_1=\varepsilon_i$, $R_2=X_i$, $R_3=\Omega$, and $R_4=w_i$, parts (i) and (iii) of the lemma imply $\{\varepsilon_i \amalg w_i \mid X_i, \varepsilon_i \amalg X_i\}$ iff $\{\varepsilon_i \amalg X_i \mid w_i, \varepsilon_i \amalg w_i\}$. Since $\varepsilon_i \amalg X_i$ is maintained, $\varepsilon_i \amalg w_i \mid X_i$ implies $\varepsilon_i \amalg X_i \mid w_i$ and $\varepsilon_i \amalg w_i$, while $\varepsilon_i \not\amalg w_i \mid X_i$ implies either $\varepsilon_i \not\amalg X_i \mid w_i$ or $\varepsilon_i \not\amalg w_i$ or both. To complete the proposition, we need to rule out $\varepsilon_i \not\amalg w_i$ with $\varepsilon_i \amalg X_i \mid w_i$ when $\varepsilon_i \not\amalg w_i \mid X_i$. Then $\varepsilon_i \not\amalg w_i \mid X_i$ always implies $\varepsilon_i \not\amalg X_i \mid w_i$. Note that

$$0 = E[\varepsilon_i \mid X_i] = E[E(\varepsilon_i \mid X_i, w_i) \mid X_i]$$

If $\varepsilon_i \not\amalg w_i$ with $\varepsilon_i \amalg X_i \mid w_i$, this implies

$$0 = E[E(\varepsilon_i \mid w_i) \mid X_i] = E(\varepsilon_i \mid w_i=0) + [E(\varepsilon_i \mid w_i=1)-E(\varepsilon_i \mid w_i=0)]P[w_i=1 \mid X_i].$$

But this cannot be true when $P[w_i=1 \mid X_i]$ takes on 2 or more distinct values since $E(\varepsilon_i \mid w_i)$ is not zero and $E(\varepsilon_i \mid w_i=1)$ is not equal to $E(\varepsilon_i \mid w_i=0)$.

The property $\varepsilon_i \amalg X_i \mid w_i$ means that estimators which are based on Assumption 1 will be consistent or unbiased in the selected sample as well as in a random sample. The property $w_i \amalg \varepsilon_i \mid X_i$ means that selection status is independent of the regression error conditional on X_i. Thus, any selection mechanism that is related to the dependent variable conditional on the regressors causes Assumption 1 to fail in the selected sample if selection status is also a function of X_i. Conversely, if Assumption 1 fails in the selected sample, selection status must be related to the dependent variable. Of course, the first point is well-known for latent-index models with homoscedastic errors (in which case, $w_i \amalg \varepsilon_i \mid X_i$ means that ε_i and the latent-index error η_i are uncorrelated). The proposition shows that this and the converse are general features of sample selection problems. In particular, these properties hold regardless of the underlying selection mechanism or model.

Proposition 1 applies equally to experimental data. For illustration, assume constant treatment effects where $y_{i1}=y_{i0}+\beta$ and $y_{i0}=\mu+\varepsilon_i$ denote the counterfactual outcomes for treated and nontreated individuals. Let X_i be an indicator of randomized treatment assignment. By virtue of random assignment, a simple comparison of means provides an unbiased estimator of β in random samples from the population at risk of treatment. But the independence between randomized treatment assignment and counterfactual outcomes is lost whenever the experiment is analyzed conditional on a variable that is both correlated with outcomes and affected by the treatment assignment. Suppose that w_i is a second outcome variable in the experiment, correlated for with y_i for some reason other than X_i. Examples of such correlated outcome pairs include death and health status, or employment status and wages. Correlation between w_i and ε_i given X_i means that $\varepsilon_i \not\perp\!\!\!\perp w_i \mid X_i$. If, as seems likely, w_i is also affected by X_i, then Proposition 1 implies that estimation of β conditional on w_i is necessarily subject to selection bias.[5]

Finally, note that Proposition 1 is also relevant for IV estimation. For applications involving instrumental variables, the key identifying assumption is that the instruments (playing the role of X_i) are independent of the regression error term (defined around an endogenous regressor) in the target population. The implications of Proposition 1 for IV estimation are that in cases where the sample selection rule involves the instruments, the instruments will be independent of ε_i in the selected sample if and only if selection status is independent of ε_i conditional on the instruments.

[5]Rosenbaum (1984) makes a similar point regarding the consequences of conditioning on "post-treatment variables" in experiments.

3.2 The selection propensity score

The *selection propensity score* is the conditional probability of selection given X_i. The fact that the selection propensity score is a function of X_i is emphasized by the notation

$$e(X_i) = P(w_i \mid X_i),$$

which is used in the definition below:

Definition 2. The selection propensity score is said to control selection bias if $\varepsilon_i \amalg X_i \mid w_i, e(X_i)$.

If conditioning on the selection propensity score controls selection bias, it may be possible to estimate β in the selected sample because then we have

$$E[y_i \mid X_i, w_i=1] = X_i'\beta + E[\varepsilon_i \mid w_i=1, e(X_i)]. \quad (4)$$

Of course, for this to be of any practical use, that must be enough variability in X_i to identify β once $e(X_i)$ is fixed. In the literature on semi-parametric selection models, this variation is obtained by imposing *a priori* exclusion restrictions on β and/or on $e(X_i)$. See Chamberlain (1986) for a precise statement of the restrictions required for identification in this context. Here I focus on Definition 2 as a likely starting point for any possible strategy to identify β.

The relationship between Definition 2 and other the conditional independence properties of sample selection models is outlined in the following proposition:

Proposition 2. The following are equivalent:

(i) $w_i \amalg X_i \mid \varepsilon_i, e(X_i)$

(ii) $(\varepsilon_i, w_i) \amalg X_i \mid e(X_i)$

(iii) $\varepsilon_i \amalg X_i \mid w_i, e(X_i)$

Proof. First observe that $w_i \amalg X_i \mid e(X_i)$ because $P[w_i=1 \mid X_i, e(X_i)] = e(X_i)$ and that $\varepsilon_i \amalg X_i \mid e(X_i)$

because $\varepsilon_i \amalg X_i$ and $e(X_i)$ is a function of X_i. Moreover, lemma 1 implies:

$$\{(i), \varepsilon_i \amalg X_i \mid e(X_i)\} \Leftrightarrow (ii) \Leftrightarrow \{(iii), w_i \amalg X_i \mid e(X_i)\}. \quad (6)$$

Therefore, since $w_i \amalg X_i \mid e(X_i)$ and $\varepsilon_i \amalg X_i \mid e(X_i)$ hold automatically in this case, the result is established.[6]

Line (iii) of Proposition 2 is the same as Definition 2, i.e., this part of the proposition consists of the statement that conditioning on the selection propensity score controls selection bias. Lines (i) and (ii) of the proposition therefore provide equivalent necessary and sufficient conditions for Definition 2. Note that while the statements $w_i \amalg X_i \mid e(X_i)$ and $\varepsilon_i \amalg X_i \mid e(X_i)$ automatically hold, line (ii) makes the stronger statement that (ε_i, w_i) are *jointly* independent of X_i given $e(X_i)$. Of course, this joint independence is not implied by marginal independence. Therefore, one consequence of the if-and-only-if relationship between (ii) and (iii) is that conditioning on the selection propensity score does not *necessarily* control selection bias.

The equivalence of lines (i) and (iii) also has practical consequences because line (i) is easy to verify in the context of specific models for the mechanism determining selection status. In particular, a weak sufficient condition for (i) is given in Proposition 3:

Proposition 3. Suppose the selection mechanism is *monotonic*, i.e., for any two values x^1, x^0 in the support of X_i, either

$$\begin{aligned} &P[w_i=1 \mid x^1, \varepsilon_i] \geq P[w_i=1 \mid x^0, \varepsilon_i], \text{ a.s. or} \\ &P[w_i=1 \mid x^1, \varepsilon_i] \leq P[w_i=1 \mid x^0, \varepsilon_i], \text{ a.s.,} \end{aligned} \quad (7)$$

where conditioning on ε_i is what makes these inequalities random. Then conditioning on the probability of selection controls selection bias.

---

[6]Choi (1992) shows that part (ii) of Proposition 2 [joint independence], imposed on a latent-index error and the regression error [i.e., replacing w_i with η_i in (2)], implies part (iii).

Proof. Note that if $e(x^1)=e(x^0)$,

$$0 = \int \{P[w_i=1 \mid x^1, e(x^1), \varepsilon_i]-P[w_i=1 \mid x^0, e(x^1), \varepsilon_i]\}h(\varepsilon_i)d\varepsilon_i.$$

Given a monotonic selection mechanism, the quantity in brackets is always non-negative. For the integral to equal zero, we must therefore have $P[w_i=1 \mid x^1, e(x^1), \varepsilon_i]=P[w_i=1 \mid x^0, e(x^1), \varepsilon_i]$. This argument shows that for a monotonic selection mechanism,

$$P[w_i=1 \mid x^1, \varepsilon_i]=P[w_i=1 \mid x^0, \varepsilon_i] \text{ whenever } e(x^1)=e(x^0). \tag{8}$$

The proposition can then be established by observing that (8) implies line (i) of Proposition 2, i.e., that the probability of selection is conditionally independent of X_i given $e(X_i)$ and ε_i.

To get an idea of how general this result is, note that any latent-index selection mechanism with constant coefficients and errors independent of X_i is monotonic. This fact does not depend on other assumptions about the error distribution. Monotonicity is satisfied by a wide range of less restrictive models as well. Consider, for example, a latent index model with random coefficients:

$$w_i = 1[X_i'\delta_i - \eta_i > 0],$$

which can be re-written,

$$w_i = 1[X_i'\delta^* - \eta_i^* > 0],$$

where $\delta^* = E[\delta_i]$ and $\eta_i^* = \eta_i + X_i(\delta^*-\delta_i)$ is an error term that is not independent of X_i even if η_i is independent of X_i. As long as the coefficient δ_i has the same sign for all i, Proposition 3 implies that conditioning on the selection propensity score controls selection bias in this model.

Finally, it is worth emphasizing the role played by full independence in obtaining the result that conditioning on the selection propensity score controls selection bias. Suppose that instead of (1) we have a heteroscedastic regression model,

$$y_i = X_i'\beta + [X_i'\gamma]\varepsilon_i \equiv X_i'\beta + \xi_i, \tag{9}$$

where ε_i is still assumed independent of X_i. The heteroscedastic error term, ξ_i, is only mean-independent of X_i. In this case, conditioning on the probability of selection does not control selection bias because

$$E[y_i | X_i, w_i=1] = X_i'\beta + E[\xi_i | w_i=1, X_i]$$

$$= X_i'\beta + [X_i'\gamma]E[\varepsilon_i | w_i=1, e(X_i)]. \qquad (10)$$

Equation (10) means that $E[\xi_i | w_i=1, X_i]$ is a function of X_i and not just $e(X_i)$. In other words,

$$\xi_i \not\perp\!\!\!\perp X_i | w_i, e(X_i).$$

Even if $E[\varepsilon_i | w_i=1, e(X_i)]$ is fixed at a number $\tilde{e}$, estimates in the selected sample are biased and converge to $\beta+\gamma\tilde{e}$. The identification failure occurs in this case in spite of the fact that the selection mechanism still satisfies part (i) of Proposition 2, i.e., $w_i \perp\!\!\!\perp X_i | \xi_i, e(X_i)$.

4. Conclusions

This paper lays out some of the basic conditional independence properties of sample selection models, focusing on those that justify conditioning on the selection propensity score to control selection bias. One reason I think this approach to sample selection models is valuable is that it leads to a very general formulation of key identifying assumptions. Another is that it leads to a weak sufficient condition for the validity of identification strategies that either explicitly or implicitly involve conditioning on the selection propensity score. Finally, the results given here are cast in terms of potentially observable quantities (w_i, X_i, and ε_i), and not inherently unobservable latent index error terms. This means that researchers could in principle construct an experiment that would test the identifying assumptions.

An additional contribution of this paper may be to help bridge the gap between the approaches taken by econometricians' and statisticians' to the problem of selection bias. Use of the propensity score to control confounding in evaluation research was introduced by Rosenbaum and Rubin (1983) and many statisticians are familiar with this idea. In a discussion of econometric models for program evaluation, Holland (1989, page 876) notes that econometricians also use the propensity score, but he suggests that this use is "quite different for the two approaches." Similarly, Heckman and Robb (1986) argue that the econometric approach to selection models uses the propensity score "in a different way than that advocated by Rosenbaum and Rubin [1983]." This paper shows that differences between the role played by the propensity score in the econometrics and statistics literature may not be as great as previously believed. In evaluation research, conditioning on the propensity score controls for the bias caused by an association between *treatment status* and observed exogenous covariates, while in selection models, conditioning on the propensity score controls for the bias caused by an association between *selection status* and observed exogenous covariates. In both cases, conditioning on the propensity score solves a problem of confounding that arises because of dependence between a dummy endogenous regressor and other covariates.

References

Ahn, H., and J.L. Powell (1993), "Semi-parametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics* 58, 3-29.

Angrist, J.D. (1995), "Conditioning on the Probability of Selection to Control Selection Bias," NBER Technical Working Paper No. 181, June.

Chamberlain, G.C. (1982), "The General Equivalence of Granger and Sims Causality," *Econometrica* 50, 569-581.

Chamberlain, G. C. (1986), "Asymptotic Efficiency in Semi-parametric Models with Censoring," *Journal of Econometrics* 32, 198-218.

Choi, Kyungsoo (1992), "Semiparametric Estimation of Sample Selection Models Using a Series Expansion and the Propensity Score," University of Chicago Department of Economics, mimeo.

Dawid, A.P. (1979), "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society Series B* 41, 1-31.

Florens, J.P, and M. Mouchart (1982),"A Note on Noncausality," *Econometrica* 50, p. 583-591.

Gronau, R. (1974), "Wage Comparisons, A Selectivity Bias," *Journal of Political Economy* 82, 1119-1144.

Ham, John, and R.J. Lalonde (1996), "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data," *Econometrica* 64 (January), 175-205.

Hartigan, John, and J. Tukey (1986), "Discussion 3: Comments on 'Alternative Methods for Evaluating the Impact of Interventions," by J. Heckman and R. Robb, in H. Wainer, ed., *Drawing Inferences from Self-selected Samples*, New York: Springer-Verlag, pp. 57-62.

Heckman, J.J., (1974), "Shadow Prices, Market Wages, and Labor Supply," *Econometrica* 42, 679-693.

Heckman, J.J., (1976), "The Common Structure of Statistical Models of Truncation, Sample selection, Limited Dependent variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement* 5, 475-492.

Heckman, J.J., (1979), "Sample Selection Bias as a Specification Error," *Econometrica* 47, 153-161.

Heckman, J.J., (1990), "Varieties of Selection Bias," *American Economic Review* 80, 313-318.

Heckman, J.J., and Richard Robb (1986), "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in H. Wainer, ed., *Drawing Inferences from Self-selected Samples*, New York: Springer-Verlag.

Holland, P. (1989), "It's Very Clear," Comment on J.J. Heckman and V.J. Hotz, "Choosing Among Alternative methods of Estimating the Impact of Social programs: The Case of Manpower Training," *Journal of the American Statistical Association* 84, 875-877.

Lee, Lung-Fei (1982), "Some Approaches to the Correction of Selectivity Bias," *Review of Economic Studies* 49, 355-372.

Little, Roderick J.A., "A Note About Models for Selectivity Bias," *Econometrica* 53, 1469-1488.

Mroz, T.A. (1987), "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica* 55, 765-799.

Newey, W.K, J.L. Powell, and J. Walker (1990), "Semi-parametric Estimation of Selection Models: Some Empirical Results," *American Economic Review* 80(2), 324-328.

Newey, W.K. (1988), "Two-Step Series Estimation of Sample Selection Models," mimeo, Department of Economics, Princeton University, October.

Olsen, R.J., (1980), "A Least Squares Correction for Selectivity Bias," *Econometrica* 48, 1815-1820.

Olsen, R.J., (1982), "Distributional Tests for Selectivity Bias and a More Robust Likelihood Estimator," *International Economic Review* 23, 223-240.

Powell, B., and L.C. Steelman (1984), "Variations in State SAT Performance: Meaningful or Misleading?," *Harvard Education Review* 54, 389-412.

Robinson, P.M., (1988), "Root-N-Consistent Semi-parametric Regression," *Econometrica* 56, 931-954.

Rosenbaum, P.R., (1984), "The Consequences of Adjustment for a Concomitant Variable that has Been Affected by the Treatment," *Journal of the Royal Statistical Society Series A* 147 Part 5, 656-666.

Rosenbaum, P.R., and D.B. Rubin, (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70, 41-55.

Rosenbaum, P.R., and D.B. Rubin, (1984), "Reducing Bias in Observational studies Using Subclassification on the Propensity Score," *JASA* 79, 516-524.

Rosenbaum, P.R., and D.B. Rubin, (1985), "Constructing a Control Group using Multi-variate Matching Methods that include the Propensity Score," *American Statistician* 39, 33-38.

Wainer, H (1986), "The SAT as a Social Indicator: A Pretty Bad idea," in H. Wainer, ed., *Drawing Inferences from Self-selected Samples*, New York: Springer-Verlag.