TECHNICAL WORKING PAPER SERIES

CONDITIONING ON THE PROBABILITY OF SELECTION TO CONTROL SELECTION BIAS

Joshua D. Angrist

Technical Working Paper No. 181

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1995

This paper was written while the author was visiting the MIT Department of Economics. Thanks go to Jinyong Hahn, Guido Imbens, Whitney Newey, Randy Olsen, and Tom Stoker for a number of helpful discussions and comments. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #181
June 1995

CONDITIONING ON THE PROBABILITY OF SELECTION TO CONTROL SELECTION BIAS

ABSTRACT

Problems of sample selection arise in the analysis of both experimental and non-experimental data. In clinical trials to evaluate the impact of an intervention on health and mortality, treatment assignment is typically nonrandom in a sample of survivors even if the original assignment is random. Similarly, randomized training interventions like National Supported Work (NSW) are not necessarily randomly assigned in the sample of working men. A non-experimental version of this problem involves the use of instrumental variables (IV) to estimate behavioral relationships. A sample selection rule that is related to the instruments can induce correlation between the instruments and unobserved outcomes, possibly invalidating the use of conventional IV techniques in the selected sample. This paper shows that conditioning on the probability of selection given the instruments can provide a solution to the selection problem as long as the relationship between instruments and selection status satisfies a simple monotonicity condition. A latent index structure is not required for this result, which is motivated as an extension of earlier work on the propensity score. The conditioning approach to selection problems is illustrated using instrumental variables techniques to estimate the returns to schooling in a sample with positive earnings.

Joshua D. Angrist
Department of Economics
Hebrew University
Mt. Scopus
Jerusalem 91905
ISRAEL
and NBER

1. Introduction

Problems of sample selection are common in both randomized and observational studies. Selection problems arise in clinical trials when a single randomized trial is used to estimate treatment-control differences in health outcomes among surviving members of the study sample. Suppose, for example, that the treatment under study keeps relatively sick patients alive while doing little for those who are otherwise healthy. This sort of impact induces a spurious negative correlation between treatment status and health status in the sample of survivors. Therefore, even if the treatment is randomly assigned at the beginning of an experimental study, it should not be viewed are randomly assigned among survivors (Kleinbaum, Kupper, and Morgenstern, 1982). An econometric example of this problem has recently been discussed by Ham and Lalonde (1995). They note that the randomized treatment (manpower training) in the experimental NSW demonstration raised trainees' employment rates as well as their average earnings. Therefore the treatment cannot be assumed to be randomly assigned in the sample of working men.

A similar problem occurs in a non-experimental context when instrumental variables (IV) techniques are used to estimate causal effects. An example is the use of IV to estimate the effect of schooling on earnings. Angrist and Krueger (1991, 1992) have shown that someone born later in the year is more likely than someone born earlier in the year to have been kept in school by compulsory attendance laws. If quarter of birth affects earnings through schooling alone, then quarter of birth dummies provide valid instruments for the relationship between schooling and earnings in the population of all men. The selection problem in this case is that schooling, and hence quarter of birth, affects the probability of working as well as the level of earnings conditional on working. A relationship between quarter of birth and employment status can therefore induce correlation between quarter of birth and errors in the wage equation in the population of working men.

Selection problems are also common in labor-supply models of the relationship between wages and the observed characteristics of workers. Because the same covariates affect wages and the probability of working, wage equations in a sample with positive earnings are not necessarily informative about the relationship between observed characteristics and offered wages (see, e.g., Heckman, 1974; Gronau, 1974). Finally, a similar problem arises in estimates of labor supply functions. Mroz (1987) reports a range of instrumental variables estimates of labor supply equations computed using methods that take the process of sample selection (conditioning on positive hours worked) into account.

One solution to such selection problems involves an assumption of joint normality for unobservables and the exclusion of some observed regressors from either the outcome equation (e.g., wages) or the sample selection rule. This leads to a selection correction in the form of inverse Mills-ratio terms (ratios of normal densities to normal CDFs) added to the outcome regression function (see Heckman, 1976 and 1979). A number of less restrictive semi-parametric variations on this normal selection model have also been introduced (see Newey, Powell, and Walker, 1990 and Heckman, 1990 for references.) A common thread in all these formulations, however, is the use of a latent index framework to characterize the mean of unobserved regression error terms conditional on the regression covariates and the sample selection rule.

An alternate approach to the selection problem has recently been discussed by Ahn and Powell (1993), who begin with the observation (noted earlier by Heckman and Robb, 1986, and Choi, 1992) that in latent index models, the selected mean of the regression error is an invertible function of the probability of selection given covariates. Ahn and Powell use this property to eliminate selection bias by differencing observations with the same (or similar) probabilities of selection. The Ahn and Powell approach sidesteps the problem of estimating an unknown conditional mean function but justifies the nonparametric estimation strategy using traditional single-index selection models.

This paper develops an approach to the selection problem that is closely related to the Ahn and Powell (1993) strategy and leads to similar estimators. A key difference is that the approach taken here is motivated as a variant on the propensity score method developed by Rosenbaum and Rubin (1983, 1984) to control for bias that arises in the estimation of treatment effects when treatment is assigned on the basis of observed covariates. Rosenbaum and Rubin show that conditional on the propensity score -- the probability of treatment assignment given covariates -- treatment assignment is independent of the covariates. Therefore, conditioning on the propensity score fully controls for confounding by these covariates in estimates of treatment effects.

The results presented here complement both Ahn and Powell (1993) and Rosenbaum and Rubin (1983) by providing weak sufficient conditions for conditioning on the selection propensity score to control bias in a general selection problem involving instrumental variables. In particular, this paper provides conditions which imply that the joint distribution of regression errors and selection status is independent of covariates conditional on the selection propensity score. These conditions are automatically satisfied by any latent index formulation of a model with selection on unobserved characteristics (error terms), but such formulations are more restrictive than necessary.[1] The principal identifying assumption used here is that the instruments satisfy a simple monotonicity condition similar to the one described by Imbens and Angrist (1994) for evaluation models.

The paper is organized as follows: Section 2 discusses a latent index approach to the problem of selection bias in the Angrist and Krueger (1991) returns-to-schooling application. Section 3 develops an alternative approach based in the selection propensity score. Section 4 uses the Angrist and Krueger (1991) data to illustrate this approach and Section 5 concludes.

2. Selection in index models with an endogenous regressor

The equation of interest in the schooling/earnings example is:

$$(1) \qquad y_i = \beta_0 + s_i\beta_1 + \epsilon_i,$$

where y_i is the log weekly wage for workers, and s_i is the highest grade completed. The instrument, Z_i, (quarter of birth) is assumed to be independent of ϵ_i in the population of all men. To focus attention on the selection problem, the regression is taken to be linear with constant coefficients. Discussion of exogenous covariates is also postponed.

The problem with attempts to estimate (1) using IV is that y_i is observed only for men who are working. Because the probability of working is also affected by schooling, conditioning on observed births is likely to induce correlation between ϵ_i and Z_i in the

[1]Ahn and Powell (1993), Choi (1992), and Newey (1988) all develop non-parametric estimators for selection models by assuming that latent index error terms are independent of covariates conditional on the selection propensity score.

population of men who work.[2] This correlation can be characterized in a latent index model as follows. Let w_i indicate selection status (i.e., $w_i=1$ indicates working) and suppose w_i is determined by a latent index capturing the possibility that schooling affects the probability of working. In particular,

$$w_i = 1[s_i\delta_1 - \nu_i > 0]$$

where ν_i is a homoscedastic error term that could be correlated with ϵ_i but is assumed to be independent of Z_i. Suppose also that there is a first stage relationship for s_i such that:

$$s_i = \gamma_0 + Z_i\gamma_1 + \eta_i,$$

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

$$E[\epsilon_i| Z_i, w_i=1] = E[\epsilon_i| Z_i, (\gamma_0\delta_1 + Z_i\gamma_1\delta_1 > \nu_i - \eta_i\delta_1)] \neq 0,$$

even though $E[\epsilon_i| Z_i] = 0$. More importantly, it is clear from the previous expression that $E[\epsilon_i| Z_i, w_i=1]$ is, in general, a function of Z_i. Therefore Z_i is no longer a valid instrument for earnings in the selected sample of workers.

The ideal solution to this selection problem would be to find an instrument that is randomly assigned or exogenous in the population where $w_i=1$. Failing this, one econometric solution to selection problems of this type is to assume that the error terms (ν_i, η_i, ϵ_i) are jointly normally distributed, as well as homoscedastic and independent of Z_i. In that case,

(2) $$E[\epsilon_i| Z_i, (\gamma_0\delta_1 + Z_i\gamma_1\delta_1 > \nu_i - \eta_i\delta_1)] = E[\epsilon_i| \gamma_0\delta_1 + Z_i\gamma_1\delta_1 > \nu_i - \eta_i\delta_1]$$

[2]Ashenfelter and Ham (1979) and Mincer (1991) document a strong association between schooling and employment probabilities.

$$= \rho E[\nu_i - \eta_i\delta_1 | \gamma_0\delta_1 + Z_i\gamma_1\delta_1 > \nu_i - \eta_i\delta_1]$$

where ρ is the coefficient from a regression of ϵ_i on $\nu_i - \eta_i\delta_1$. Joint independence of (ν_i, η_i, ϵ_i) with Z_i allows the simplification in the first line of (2). Because of normality we can simplify further:

$$E[\nu_i - \eta_i\delta_1 | \gamma_0\delta_1 + Z_i\gamma_1\delta_1 > \nu_i - \eta_i\delta_1] = -\phi(\gamma_0\delta_1 + Z_i\gamma_1\delta_1)/\Phi(\gamma_0\delta_1 + Z_i\gamma_1\delta_1)$$

$$\equiv \lambda(\gamma_0\delta_1 + Z_i\gamma_1\delta_1),$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the normal density and distribution functions for $\nu_i - \eta_i\delta_1$. Adding $\lambda(\gamma_0\delta_1 + Z_i\gamma_1\delta_1)$ as a regressor to (1) provides a practical solution to the selection problem in this context because parameters in $\lambda(\gamma_0\delta_1 + Z_i\gamma_1\delta_1)$ can be estimated from a Probit regression of w_i on Z_i.

Critics of econometric selection models (e.g., Hartigan and Tukey, 1986; Little, 1985) point out that such models combine a variety of independence and functional form assumptions that can be hard to assess and interpret. The validity of parametric corrections for sample selection bias clearly turns to a large extent on distributional and functional form assumptions and the validity of exclusion restrictions (Olsen 1980, 1982). On the other hand, while the need for exclusion restrictions seems hard to avoid (unless identification is to be based solely on non-linearities induced by distributional assumptions) many authors have noted that normality is not an essential feature of the index-model approach to selection problems. See Newey, Powell, and Walker (1990) and Lee (1982) for examples.

One feature common to almost all parametric and non-parametric selection corrections is a latent index structure similar to the normal formulation, and an attempt to estimate or approximate the unknown conditional mean function, $E[\epsilon_i | \gamma_0\delta_1 + Z_i\gamma_1\delta_1 > \nu_i - \eta_i\delta_1]$.[3] An alternative approach begins with the observation that the selection-correction term is a potentially non-linear but invertible transformation of the selection propensity score, $P[w_i=1 | Z_i]$.[4] Given independence of error terms and instruments, an implication of equation (2) is that conditional on $P[w_i=1 | Z_i]$ being fixed, selection bias does not affect instrumental variables estimates of the slope parameter in (1). This fact motivates the semi-parametric estimators discussed by Ahn and Powell (1993) and Choi (1992) for models with exogenous regressors. In particular, Ahn and Powell suggested that β_1 be estimated by differencing observations for which non-parametric estimates of $P[w_i | Z_i]$ are close. The next section discusses weak sufficient conditions for the conditional independence property underlying the Ahn and Powell and related estimation strategies to hold.

3. Conditioning on the selection propensity score

The structure retained from the previous section is the outcome equation, (1), maintaining the assumption that ϵ_i is independent of the instrument. Now let the instrument be a discrete scalar variable, Z_i, taking on one of J values as follows:

$$Z_i \in \{z_1, z_2, \ldots, z_J\}.$$

Note that an implication of equation (1) is that:

---

[3]An exception is Manski (1994), who develops bounds for selection bias in more general models.

[4]Heckman and Robb (1986, p. 102-104) discuss this point in the context of models involving selection on observables and in the context of linear latent index models with selection on unobservables.

(3) $E[y_i | Z_i, w_i=1] = \beta_0 + E[s_i | Z_i, w_i=1]\beta_1 + E[\epsilon_i | Z_i, w_i=1]$.

Conditioning on any random variable, X_i, can help identify β_1 if there are two possible realizations of Z_i, denoted z_j and z_k, such that the following holds:

Condition 1. $E[\epsilon_i | Z_i=z_j, w_i=1, X_i] = E[\epsilon_i | Z_i=z_k, w_i=1, X_i]$.

The reason Condition 1 is useful is that as long as

$$E[s_i | Z_i=z_j, w_i=1, X_i] \neq E[s_i | Z_i=z_k, w_i=1, X_i],$$

β_1 is identified by comparisons of the mean of y_i at $Z_i=z_j$ and $Z_i=z_k$ in the selected sample. More generally, the instruments can be used to estimate β_1 by two-stage least squares (2SLS) in samples where X_i is fixed.

Ahn and Powell (1993) show that Condition 1 is also necessary for nonparametric identification of β_1 in the following sense.[5] Equation (3) implies that equation (1) can be written

(4a) $y_i = \beta_0 + s_i\beta_1 + \lambda(Z_i) + \mu_i$,

where $\lambda(Z_i) \equiv E[\epsilon_i | Z_i, w_i=1]$ and μ_i is an error term that satisfies $E[\mu_i | Z_i, w_i=1]=0$ by construction. In the absence of distributional assumptions on ϵ_i other than independence of Z_i, this is the only implication of equation (1). Note that if $\lambda(Z_i)$ is completely unrestricted, the following regression also holds:

(4b) $y_i = \beta_0 + s_i\beta_1^* + \lambda^*(Z_i) + \mu_i^*$,

where $\beta_1^* = \beta_1 + \alpha$, $\lambda^*(Z_i) = \lambda(Z_i)\text{-}E[s_i | Z_i, w_i=1]\alpha$, and $\mu_i^* = \mu_i + (E[s_i | Z_i, w_i=1]\text{-}s_i)\alpha$.

---

[5]Ahn and Powell's discussion of this point is for models with exogenous regressors.

Therefore, because $E[\mu_i^* | Z_i, w_i=1]=E[\mu_i | Z_i, w_i=1]=0$, β_1 cannot be distinguished from β_1^* without further restrictions.[6]

3.1 Assumptions and results

This section establishes a basic conditional independence result which implies that Condition 1 holds when X_i is the selection propensity score. At this point, it is useful to introduce new notation that describes how the sample selection rule is related to the instruments. Define indicators of *potential selection status* to be a set of J random variables, w_{ij}, each of which describes the selection status of observation i when $Z_i = z_j$. I assume that a full set of w_{ij} could be revealed for each individual, either by careful questioning or by experimentation, even though only one is actually observed in practice. To illustrate the nature of potential selection status, note that in the latent index formulation of the previous section, each w_{ij} is given by:

$$w_{ij} = 1[\gamma_0\delta_1 + z_j\gamma_1\delta_1 > \nu_i - \eta_i\delta_1];\ j = 1, \ldots, J.$$

More generally, potential selection status is analogous to the notion of potential treatment assignment used by Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1995) to

[6]Chamberlain (1986) derives the restrictions necessary and sufficient for there to be positive information about β_1 in a model where selection is determined by a linear latent index. Chamberlain also argues that for practical purposes, conditions for positive information should be interpreted as conditions for identification. These conditions can be summarized as requiring either exclusions restrictions on the selection propensity score, or the combination of exclusion restrictions in the list of regressors and a continuously distributed selection propensity score. The former seems more natural in an IV context.

characterize IV estimates of treatment effects.[7]

The relationship between potential treatment status, potential outcomes, and the instruments is restricted as follows:

Assumption 1 (Independence). $\{\epsilon_i, (w_{ij}; j = 1, 2, \ldots, J)\}$ are jointly independent of Z_i.

An implication of Assumption 1 is that the marginal distribution of each w_{ij} is identified from a sample with data on Z_i and w_i. The joint distribution of $(w_{ij}; j = 1, 2, \ldots, J)$ is not identified from observational data because only one w_{ij} is observed for any one person.

The latent index formulation clearly satisfies Assumption 1 because the selection index error term, $\nu_i - \eta_i\delta_1$, is independent of the instruments. But the latent index structure is not necessary for Assumption 1 to hold. For example, Assumption 1 is plausible in an application where the instruments are generated by random assignment or where they can be thought of as constituting a "natural experiment". For example, if quarter of birth is assigned (by nature and parents) without regard to earnings potential or to how likely a man born in different quarters would be to work, then Assumption 1 is satisfied for the Angrist and Krueger (1991) application.

Potential selection status does not have to be tied to an underlying parametric model for meaningful and potentially verifiable assumptions about treatment assignment to be described. In particular, the only assumption required beyond Assumption 1 restricts the

[7]The notation for potential treatment assignment was originally suggested by Gary Chamberlain in private correspondence with Angrist and Imbens.

relationship between the instruments and potential selection status to be uni-directional, or monotone:

Assumption 2 (Monotonicity). Two indicators of potential treatment status, w_{ij} and w_{ik}, are said to satisfy monotonicity if either $w_{ij} \geq w_{ik}$ for all i or $w_{ij} \leq w_{ik}$ for all i.

Monotonicity is automatically satisfied by w_{ij} and w_{ik} when they are determined by a linear latent index model because for all j and k:

$$w_{ij} = 1[\gamma_0\delta_1 + z_j\gamma_1\delta_1 > \nu_i - \eta_i\delta_1] \quad \text{and} \quad w_{ik} = 1[\gamma_0\delta_1 + z_k\gamma_1\delta_1 > \nu_i - \eta_i\delta_1].$$

Here it is clearly true that if $w_{ij} > w_{ik}$ for any i, then $w_{ij} \geq w_{ik}$ for all i, and likewise for the reverse inequality. More general latent index models also share the monotonicity property. Suppose $w_{ij} = 1[f(z_j) > \nu_i^*]$ for any function $f(\cdot)$ and any random variable ν_i^* that is independent of Z_i.[8] The resulting sequence of w_{ij} clearly satisfies both Assumptions 1 and 2. The next section shows that many non-index models have this monotonicity property as well.

An important difference between Assumption 2 and the index model formulation (which imposes Assumption 2 by default) is that the index model casts identifying assumptions as restrictions on latent indices that have no empirical counterpart (because observed quantities are a non-invertible function of the latent error terms.) In contrast, it is possible to imagine designing a survey that would allow Assumption 2 to be checked. For example, we could ask individuals how they would act or what they think would happen to

---

[8]Ahn and Powell (1993) refer to selection indicators of this type as having the "single-index property."

them if they were exposed to different values of Z_i. In cases where Z_i can be manipulated and w_{ij} is time-invariant for all j, we could actually test Assumption 2 by exposing individuals repeatedly to alternative values of Z_i in a controlled experiment.[9]

To assess the plausibility of monotonicity without actually conducting a separate study a researcher must describe (model) the manner in which the instruments affect selection. This model can come from an understanding of what sort of experiment the instrument indicates in a clinical setting, or is meant to replicate in an observational study. For example, in a clinical trial with selection by survivorship, a researcher may be prepared to assert that any pairwise comparison of two randomly assigned dosage levels affects survivorship in the same way (or not at all) for all study participants. This will be true if some dosages are known to be non-harmful. Note, however, that Assumption 2 does not require the relationship between dosage and survivorship to be monotone in dosage.

The theoretical importance of Assumptions 1 and 2 is that together they imply that conditional on $P(w_i=1 \mid Z_i)$, the joint distribution of $\{w_i, \epsilon_i\}$ is independent of Z_i.

Proposition 1. Given Assumptions 1 and 2, $\{w_i, \epsilon_i\} \amalg Z_i \mid P(w_i \mid Z_i)$.

Proof. Consider two values of Z_i, denoted by z_j and z_k, for which w_{ij} and w_{ik} satisfy monotonicity and for which $P(w_i=1 \mid Z_i=z_j)=P(w_i=1 \mid Z_i=z_k) \equiv e(z_j)$. To simplify notation, write $P(w_i=1 \mid Z_i=z_j, w_i=1, P(w_i=1 \mid Z_i)=e(z_j)]$ as $P[w_i=1 \mid z_j, w_i=1, e(z_j)]$. The result is

[9]Haavelmo (1944, page 6) makes a similar tie to experimentation in his interpretation of simultaneous equations models. In particular, Haavelmo argues that for a model to be meaningful it should be possible to conceive of an experiment that would reveal theoretical quantities, even though such an experiment could be difficult to carry out in practice.

established by first showing that

(5) $\quad P[w_i=1 \mid z_j, e(z_j), \epsilon_i] = P[w_i=1 \mid z_k, e(z_j), \epsilon_i].$

Let $h(\epsilon_i)$ be the marginal density of ϵ_i. Iterating expectations and using the fact that ϵ_i is independent of Z_i, we have:

$$0 = \int \{P[w_i=1 \mid z_j, e(z_j), \epsilon_i] - P[w_i=1 \mid z_k, e(z_j), \epsilon_i]\} h(\epsilon_i) d\epsilon_i$$

because $P[w_i=1 \mid z_j, e(z_j)] = e(z_j)$ and $P(w_i=1 \mid z_k, e(z_j)] = e(z_j)$. Note that Assumption 1 implies $P[w_i=1 \mid z_j, e(z_j), \epsilon_i] = P[w_{ij}=1 \mid \epsilon_i]$. Therefore,

$$0 = \int \{P[w_{ij}=1 \mid \epsilon_i] - P[w_{ik}=1 \mid \epsilon_i]\} h(\epsilon_i) d\epsilon_i.$$

By Assumption 2, $\{P[w_{ij}=1 \mid \epsilon_i] - P[w_{ik}=1 \mid \epsilon_i]\}$ must be either non-positive for all ϵ_i or non-negative for all ϵ_i. Therefore, since $h(\epsilon_i)$ is always non-negative, $P[w_{ij}=1 \mid \epsilon_i]$ must equal $P[w_{ik}=1 \mid \epsilon_i]$ for all ϵ_i and (5) must hold. The proof is completed by noting that

(6) $\quad P[w_i=1 \mid z_j, e(z_j), \epsilon_i] = P[w_i=1, \epsilon_i \mid z_j, e(z_j)]/h(\epsilon_i).$

Combined with equation (5) this implies

$$P[w_i=1, \epsilon_i \mid z_j, e(z_j)] = P[w_i=1, \epsilon_i \mid e(z_j)].$$ □

To interpret Proposition 1, note that Assumption 1 alone implies

$$\epsilon_i \perp\!\!\!\perp Z_i \mid P(w_i \mid Z_i)$$

because $\epsilon_i \perp\!\!\!\perp Z_i$. In other words, ϵ_i is independent of Z_i given the selection propensity score. Second, note that Rosenbaum and Rubin (1983) have shown that any binary variable is independent of covariates given the propensity score.[10] In the notation of this paper, we have

[10]This simple but useful result can be established in a few lines. Let d_i be an indicator of treatment assignment that is correlated with covariates, X_i, and define $e(X_i) \equiv P[d_i=1 \mid X_i]$. Note that $P[d_i=1 \mid X_i, e(X_i)] = P[d_i=1 \mid X_i] = e(X_i)$. Also, $P[d_i=1 \mid e(X_i)] = E\{E[d_i=1 \mid X_i,$

$$w_i \amalg Z_i \mid P(w_i \mid Z_i).$$

In other words, w_i is independent of Z_i given the selection propensity score. The contribution of Proposition 1 is to show that Assumptions 1 and 2 together imply that w_i and ϵ_i are also *jointly* independent of Z_i conditional on $P(w_i \mid Z_i)$. Marginal conditional independence, i.e., $w_i \amalg Z_i \mid P(w_i \mid Z_i)$, is similar to the conditional independence property that underlies matching methods using the propensity score in evaluation research. It is worth emphasizing, however, that the joint conditional independence result established in Proposition 1 requires stronger assumptions (in this case, monotonicity) than does marginal conditional independence of the Rosenbaum and Rubin sort.

The following implication of Proposition 1 is the key result used in the next section to justify estimation procedures such as the one discussed by Ahn and Powell (1993).

Corollary. Given Assumptions 1 and 2,

(7) $$E[\epsilon_i \mid Z_i=z_j,\ w_i=1,\ P(w_i=1 \mid Z_i)=e(z_j)] = E[\epsilon_i \mid Z_i=z_k,\ w_i=1,\ P(w_i=1 \mid Z_i)=e(z_j)].$$

In other words, Condition 1 is satisfied by the selection propensity score.

Proof. Let $h[\epsilon_i \mid z_j, w_i=1, e(z_j)]$ denote the conditional density of ϵ_i given $Z_i=z_j$ and $w_i=1$. The conditional independence property in Proposition 1 implies that

(8) $$h[\epsilon_i \mid z_j, w_i=1, e(z_j)] = h[\epsilon_i \mid z_k, w_i=1, e(z_j)] \equiv h[\epsilon_i \mid w_i=1, e(z_j)].$$ □

$e(X_i)] \mid e(X_i)\} = E\{e(X_i) \mid e(X_i)\} = e(X_i)$. Therefore, $P[d_i=1 \mid X_i, e(X_i)] = P[d_i=1 \mid e(X_i)]$. The importance of this result for evaluation research is that conditioning on $P(d_i \mid X_i)$ fully controls for possible confounding by X_i when estimating the effect of d_i on outcomes. See Rosenbaum and Rubin (1984) for an empirical example.

Note that one simple implication of this corollary is that, given Assumptions 1 and 2, if $P(w_i|Z_i)$ does not vary with Z_i then sample selection does not affect the consistency of IV estimates.

3.2 Monotonicity of selection status in earnings functions

In the example outlined in Section 2, monotonicity requires that quarter of birth affect the probability of working the same way for everybody. This condition can be assessed in the context of a behavioral model that links the instruments to the selection process. To illustrate this link, recall that Angrist and Krueger (1991, 1992) provide evidence which suggests that because of compulsory attendance laws, men born in late quarters get at least as much schooling as they would have gotten had they been born earlier. To formalize the idea of monotonicity in this context, let s_{ij} denote the years of schooling individual i would get if $Z_i=z_j$, for $j=1, \ldots, J$. Monotonicity of the relationship between Z_i and s_i means that for any two values z_j and z_k, either $s_{ij} \geq s_{ik}$ for all i or $s_{ij} \leq s_{ik}$ for all i. An implication of this condition is that the CDFs of schooling conditional on z_j and z_k should not cross. Angrist and Imbens (1995) show that this implication is satisfied for comparisons of schooling by quarter of birth in both the 1970 and 1980 Census.

To extend monotonicity of the relationship between Z_i and s_i to the relationship between Z_i and w_i, note that one commonly used model of the labor force participation decision postulates that individuals choose to work by comparing the potential payoff from working with some individual reservation level of earnings, r_i, which is independent of covariates (see, e.g., Heckman and Killingsworth 1986, Section 4.1). If individuals behave this way and they use equation (1) to predict what they will earn once their schooling is completed, then

(9) $$w_i = 1[y_i > r_i] = 1[\beta_0 + s_i\beta_1 + \epsilon_i > r_i] \text{ and}$$

$$w_{ij} = 1[\beta_0 + s_{ij}\beta_1 + \epsilon_i > r_i],$$

where w_{ij} is potential selection status given $Z_i = z_j$. The presumption in the schooling/earnings application is that the only reason quarter of birth affects labor market outcomes is because of schooling. In this case, $s_{ij} \geq s_{ik}$ clearly implies $w_{ij} \geq w_{ik}$. Therefore, if participation is determined by equation (9), Assumption 2 is satisfied in the Angrist and Krueger (1991) application.

It is important to note that even though (9) satisfies monotonicity it does not necessarily have the single-index property used by Ahn and Powell (1993) to justify conditioning on the selection propensity score. Suppose, for example, that $s_i = \pi_{1i}Z_i + \pi_{0i}$, where π_{0i} and π_{1i} are random coefficients distributed independently of Z_i with means π_0 and π_1. In this formulation, the sequence of potential selection indicators, w_{ij}, satisfies Assumptions 1 and 2 even though the indicators cannot be written as a function of a single index involving the instruments plus an error term that is independent of the instruments. To see this, substitute $\pi_{1i}Z_i + \pi_{0i}$ for s_i in (9):

(10) $$w_i = 1[y_i > r_i] = 1[(\beta_0 + \pi_0\beta_1) + Z_i\pi_1\beta_1 > r_i - \epsilon_i + (\pi_0 - \pi_{0i})\beta_1 + Z_i(\pi_1 - \pi_{1i})\beta_1].$$

The latent index "error term" in this case, $[r_i - \epsilon_i + (\pi_0 - \pi_{0i})\beta_1 + Z_i(\pi_1 - \pi_{1i})\beta_1]$, is necessarily dependent on Z_i.

More generally, employment status might be determined by a version of equation (10) where β_0 and β_1 are replaced by individual-specific coefficients, (β_{0i}, β_{1i}), or where $(\beta_0 + \beta_1 s_i)$

is replaced by an individual-specific non-linear function, $f_i(s_i)$. The resulting model for w_{ij} is unlikely to have the single-index property for the same reason that the formulation in equation (10) does not. But such models will satisfy monotonicity for comparisons at $Z_i=z_j$ and $Z_i=z_k$, as long as β_{1i} is positive or $f_i(s_i)$ is increasing in s_i over the range of variation induced by changing Z_i from z_j to z_k.

3.3 Estimation

An implication of the corollary to Proposition 1 is that the following non-linear regression can sometimes be used to identify β_1:

(11) $E[y_i| Z_i, w_i=1] = \beta_0 + E[s_i| Z_i, w_i=1]\beta_1 + E[\epsilon_i| w_i=1, P(w_i=1| Z_i)]$.

The formulation in Section 2 is a special case of this equation, where

$P(w_i=1| Z_i) = \Phi(\gamma_0\delta_1 + Z_i\gamma_1\delta_1)$ and

$E[\epsilon_i| w_i=1, P(w_i=1| Z_i)] = \rho\lambda(\Phi^{-1}[P(w_i=1| Z_i)])$.

Equation (11) provides a practical tool for non-parametric correction of selection bias because in many empirical applications $P(w_i=1| Z_i)$ is non-parametrically identified.[11]

The framework developed here also leads to estimators for models with exogenous covariates and models with no endogenous regressors. To see this, suppose that the instrument is also the regressor. In this case, equation (11) becomes:

(12) $E[y_i| Z_i, w_i=1] = \beta_0 + Z_i\beta_1 + E[\epsilon_i| w_i=1, P(w_i=1| Z_i)]$.

Of course, for an approach based on equation (11) or (12) to be useful, it must be the case

[11]In the context of Angrist and Krueger (1991), non-parametric identification means that it is possible to trace out the unknown $P(w_i=1| Z_i)$ by computing the population average of w_i at each value of Z_i.

that there is variation in Z_i or $E[s_i|\ Z_i,\ w_i=1]$ in subpopulations where $P(w_i=1|\ Z_i)$ is fixed. If there is such variation, then β_1 can be estimated by 2SLS or OLS in samples stratified on $P(w_i=1|Z_i)$. Alternately, an estimate of $P(w_i=1|Z_i)$ can be incorporated in the regression function.[12]

3.4 The role of exclusion restrictions

The possibility of developing a practical solution to selection problems like the one described here increases when there are instruments or regressors than can be excluded from the selection propensity score. Returning to the endogenous regressor case, suppose that Z_i is a vector with components Z_{1i} and Z_{2i}, such that

$$P(w_i=1|\ Z_i) = P(w_i=1|\ Z_{i1}).$$

Now equation (11) can be written

$$(13)\quad E[y_i|\ Z_i,\ w_i=1] = \beta_0 + E[s_i|\ Z_{i1},\ Z_{i2},\ w_i=1]\beta_1 + E[\epsilon_i|\ w_i=1,\ P(w_i=1|\ Z_{i1})].$$

In this case, β_1 can be identified by stratifying on Z_{i1}.

When equation (13) holds, estimates of β_1 can also be computed using a semi-parametric estimator that takes advantage of the fact that $P(w_i=1|Z_i)=P(w_i=1|Z_{i1})$ when estimating $E[\epsilon_i|\ w_i=1,\ P(w_i=1|\ Z_i)]$. If the additional exclusion restrictions are invalid, however, then conditioning on Z_{i1} in a highly flexible or nonparametric manner is likely to do little to reduce selection bias and could even make it worse. The notion that exclusion

---

[12]Ahn and Powell (1993) suggest that β_1 be estimated in (11) after differencing all observations for which nonparametric estimates of $P(w_i=1|\ Z_i)$ are close. A similar differencing estimator for models that control for confounding using the propensity score is discussed by Rosenbaum and Rubin (1985).

restrictions matter more than functional form is also consistent with Mroz's (1987) finding that sample selection corrections for labor supply equations are more sensitive to the choice of exclusion restrictions than to distributional assumptions.

3.5 Evaluation models

The approach outlined here is also related to the use of selection models to control for bias in applications where selection status, w_i, is actually the regressor of interest. Suppose, for example, that w_i indicates participation in a training program and that the equation of interest is:

(14) $y_i = \beta_0 + w_i\beta_1 + \epsilon_i$,

where, as before, ϵ_i is independent of Z_i but w_i is not.

Assumption 1 (independence) is clearly sufficient to identify β_1 in (14) when observations are available on a random sample from a population that includes observations where $w_i=0$ and observations where $w_i=1$, and the instruments affect w_i (see, e.g., Heckman, 1990). In fact, Assumption 1 can be weakened to require only mean independence of ϵ_i and Z_i, leading naturally to IV estimation. On the other hand, Assumption 2 and full independence as in Assumption 1 are relevant for IV estimation of (14) when the constant-coefficients framework is dropped in favor of a model with heterogeneous potential outcomes. In that case, Assumption 1 and Assumption 2 (monotonicity of the relationship between w_i and Z_i) ensures that IV estimates of (14) produces an average causal effect for those whose treatment status is influenced by the instruments (Imbens and Angrist, 1994).

Finally, note that Heckman and Robb (1986, p. 79) have pointed out that as an alternative to IV estimation, consistent estimates of β_1 can also be obtained by estimating

$$(15) \quad y_i = \beta_0 + w_i\beta_1 + [\, E(\epsilon_i | \, w_i=0, Z_i)(1-w_i) + E(\epsilon_i | \, w_i=1, Z_i)w_i \,] + \xi_i$$

by non-linear least squares in cases where the conditional mean function, $E(\epsilon_i | \, w_i, Z_i)$, is assumed to be known up to a finite set of parameters. In equation (15), $E\{\, E(\epsilon_i | \, w_i=0, Z_i)(1-w_i) + E(\epsilon_i | \, w_i=1, Z_i)w_i) \,|\, Z_i \,] = 0$ and $E[\xi_i | \, Z_i]=0$ by construction. On the other hand, absent observations where $w_i=0$, β_1 cannot be distinguished from β_0 in equation (15). Similarly, Proposition 1 does not facilitate identification of the constant in sample selection problems.

4. Sample selection corrections for IV estimates of the returns to schooling

The literature on schooling and earnings devotes considerable attention to the problem of "ability bias" in estimates of the economic returns to schooling.[13] Ability bias is a form of omitted-variables bias that would arise if more able individuals in the labor market get more schooling, perhaps because of better access to capital markets. The observed positive correlation between schooling and earnings would then partly reflect the fact that those with more schooling have higher earnings potential.

Instrumental variables that are related to earnings solely because of schooling may solve this problem. Angrist and Krueger (1991, 1992) showed that students' quarter of birth interacts with compulsory attendance laws and age at school entry to generate variation in years of completed schooling. State compulsory attendance laws typically require students to

[13]For a recent survey, see Card (1994).

enter school in the Fall of the year in which they turn six, but allow students to drop out of school when they reach their 16th birthday. This induces a relationship between quarter of birth and educational attainment because students born in the first quarter of the year enter school at an older age than students born in later quarters. Students who enter school at an older age are therefore allowed to drop out after having completed less schooling than students who enter school at a younger age. If students' quarter of birth is correlated with earnings solely because it is correlated with schooling, then it is an instrument for schooling in an earnings equation.

A set of results like those presented by Angrist and Krueger are shown in Table 1.[14] Column 3 of Table 1 shows the OLS estimate of β_1 in equation (1), when the equation is modified to include a set of year dummies as exogenous covariates. Column 4 shows the corresponding 2SLS estimate when the excluded instruments are a set of three quarter-of-birth dummies. The OLS estimate is .072 with a very small standard error and the 2SLS estimate is .110 with a standard error of .018. Columns 1 and 2 report the reduced form coefficients underlying the 2SLS estimates. Both schooling and earnings are lower for early-quarter birth cohorts than for later-quarter birth cohorts. Column 5 of Table 1 reports 2SLS estimates when the instrument list is modified to include 3 quarter-of-birth dummies for each of the 10 years of birth in the sample, generating a total of 30 instruments. Because the

[14]This sample differs from the Angrist and Krueger (1991) sample in that men with wages allocated by the Census Bureau are included in the analysis. Other sample restrictions are the same as described in the Appendix to Angrist and Krueger (1991) for the 1980 Census sample born 1930-39. Men with allocated wages are included because a substantial fraction of observations with zero wages have had wage data allocated. In the sample used here, over 34 percent of men born 1930-39 with reports of zero wages have allocated wage data, while only 11 percent of men with positive earnings have allocated wage data.

second-stage equation includes year of birth dummies, this 2SLS estimate is generated by within-year-of-birth variation in schooling and earnings by quarter of birth. The resulting estimate is .088 with a standard error of .015.

The estimates in columns 1-5 of Table 1 are for men with positive earnings in the sample born 1930-39 who satisfy the criteria described in footnote 14. In this sample, 15.3 percent have zero earnings. Column 8, which reports the results from an OLS regression of a dummy for having positive earnings on year dummies and years of schooling, shows that the probability of having positive earnings rises with schooling. In particular, additional years of schooling are associated with an average 1.2 percentage point increase in the probability of working. 2SLS estimates of the relationship between schooling and the probability of working are reported in columns 9-10. The estimates in column 9 are from a model that uses 3 quarter-of-birth dummies as instruments and the estimates in column 10 are from a model that uses 30 quarter-of-birth/year-of-birth interactions as instruments, as in column 5. Both of the 2SLS estimates are highly statistically significant and substantially larger than the corresponding OLS estimates.

The reduced form relationships underlying the 2SLS estimates reported in column 8 are shown in columns 6 and 7. The relationship between schooling and quarter of birth in column 6 differs slightly from the relationship between schooling and quarter of birth in column 1 because the sample used for column 6 includes observations with zero earnings. The estimates in column 7 show that early-quarter birth cohorts have a lower probability of working than later-quarter birth cohorts. This is probably because of the association between quarter of birth and schooling and between schooling and the probability of working.

The relationship between quarter of birth and the probability of working documented in column 7 is important because this relationship implies that IV estimates may generate biased estimates of the relationship between schooling and offered wages when computed in samples of working men. On the other hand, a simple correction strategy based on the approach outlined here can be implemented directly using the estimates in Table 1. Note that the probability of working differs little by birth quarter when the first-quarter birth cohort is compared to the second quarter birth cohort. The difference between the first and second quarter effects in the reduced form equation for w_i is .002. The difference in schooling between first-quarter and second-quarter birth cohorts is .004. Taking the probability of selection as approximately fixed in the first-quarter/second quarter contrast leads to an IV estimate of the returns to schooling that compares schooling and earnings in these two groups. The resulting calculation is $[-.015+.011]/[-.156+.106] = .08$, which is similar to the conventional 2SLS estimates.

For an alternative estimate of the same type, note that the contrast between 3rd and 4th quarters in the selection propensity score is even closer to zero than the first quarter-second quarter contrast. Estimates of the returns to schooling based on this contrast are $.002/-.037 = -.054$. Under the Assumptions required for Proposition 1, these estimates should be unaffected by sample selection bias. In fact, both estimates can be interpreted as a simple application of the Ahn and Powell (1993) semi-parametric differencing estimator. An obvious drawback of this simple approach is that none of the contrasts underlying these two illustrative calculations are statistically significant.

Results of attempts to produce more precise estimates are reported in Table 2.

Column 1 of Table 2 repeats the 2SLS estimates using 3 quarter-of-birth dummies as excluded instruments (from column 4 in Table 1). Column 2 shows the results of adding an estimate of $P[w_i=1 | Z_i]$ to this regression as a covariate.[15] In this case, the estimate of $P[w_i=1 | Z_i]$ is given by the fitted values from the reduced form regression of w_i on 9 year-of-birth dummies and 3 quarter-of-birth dummies. The quarter-of-birth coefficients from this regression are reported in column 7 of Table 1. Not surprisingly, adding the fitted values from a regression on 3 quarter dummies leads to very imprecise 2SLS estimates when the 3 dummies are also used as instruments. The resulting schooling coefficient is -.160 with a standard error of .192.

Column 3 of Table 2 repeats the 2SLS estimates (from Column 5 of Table 1) using 3 quarter-of-birth dummies for each of 10 years of birth, for a total of 30 excluded instruments. Columns 4 and 5 show the result of adding linear and quadratic terms in $P[w_i=1 | Z_i]$ as covariates to this model. In this case, the estimates of $P[w_i=1 | Z_i]$ are fitted values from a regression of w_i on 9 year of birth dummies and a full set of 30 year of birth and quarter of birth interaction terms. This regression generates a consistent non-parametric estimate of the true conditional expectation of w_i given the instruments and regressors. The results in column 3 show that adding linear and quadratic terms in the selection propensity score to the second stage equation leads to somewhat larger standard errors for schooling

[15]A similar "covariance adjustment" for selection bias is discussed by Wainer (1986) and Powell and Steelman (1984). Newey (1988) discusses the inclusion of polynomial or other expansion terms in $P[w_i | Z_i]$ as regressors in sample selection models. For single-index models, Newey shows that in the special case where the conditional expectation of regressors is linear in $P[w_i | Z_i]$, including $P[w_i | Z_i]$ as a regressor generates consistent estimates of β_1 regardless of the form of $E[\epsilon_i | w_i=1, P(w_i | Z_i)]$.

coefficients than does conventional 2sls estimation, without resulting in a dramatic change in the schooling coefficients.

Columns 6-9 of Table 2 report the results of estimating schooling coefficients using 30 quarter-of-birth/year-of-birth interactions as instruments, after stratifying on quartiles of the corresponding estimates of $P[w_i=1 | Z_i]$. The quartiles are calculated after removing year effects in the estimates of $P[w_i=1 | Z_i]$. In other words, the idea here is to hold constant most of the within-year-of-birth variation in $P[w_i=1 | Z_i]$. The within-year range (i.e., deviations from year means) in the average probability of working by quarter of birth is -.0035 to .014. The quartiles are reported in the next to last row in the Table.[16] The results of within-quartile estimation are highly variable across quartiles and imprecise, ranging from -.071 with a standard error of .207, to .167 with a standard error of .073. The average estimate across quartiles is .087 with a standard error of .06. Although this too is imprecise, it differs little from the conventional 2SLS estimate of .088 in column 5.

The fact that non-parametric procedures to control for selection bias lead to imprecise estimates is a natural consequence of the attempt to use functions of quarter of birth both as regressors and control variables. Ahn and Powell (1993, p. 20) report similar imprecision in their empirical example but note that such imprecision "offers a realistic picture of the

---

[16]This approach can be interpreted as an application of the estimator proposed by Robinson (1988). In the current context, Robinson's approach transforms a semi-parametric model of the form

$$y_i = \beta_0 + s_i\beta_1 + \lambda[e(Z_i)] + \mu_i,$$

where $\lambda[e(Z_i)]$is an unknown function of Z_i, to

$$[y_i - E(y_i | e(Z_i))] = \beta_0 + [s_i - E(s_i | e(Z_i))]\beta_1 + \mu_i.$$

The population $E(y_i | e(Z_i))$ and $E(s_i | e(Z_i))$ are then replaced with non-parametric estimates. Conditioning the estimation on quartiles of $e(Z_i)$ implements this estimator if there is no within-quartile variation in $e(Z_i)$ and β_1 is constant across quartiles.

attainable precision when the form of the selection equation is uncertain." Since the form of the selection correction is always uncertain in applied work, this imprecision would appear to be endemic in sample selection models.

Although non-parametric corrections for selection bias should probably be expected to generate inconclusive results, the fact that conventional estimates and alternative selection-correction strategies generate similar results suggests that selection bias does not have a big impact on estimates of β_1 in the earnings function studied here. For example, any role for terms involving $P(w_i | Z_i)$ in equation (1) should have been reflected in the estimates of β_1 in Table 2. The finding that corrections for selection bias matter little is also consistent with Angrist and Krueger's (1991) conclusion that there is actually very little ability bias in OLS estimates of schooling coefficients. Note that if ϵ_i and $[\nu_i - \eta_i\delta_1]$ are not actually correlated in the index-model version of the selection problem (or ϵ_i and w_{ij} in the framework of Section 3), then sample selection rules involving the instruments do not bias 2SLS estimates.

5. Conclusions

In a discussion of econometric models for program evaluation, Holland (1989, page 876) notes that both econometricians and statisticians use the propensity score in selection and evaluation models but suggests that this use is "quite different for the two approaches." Similarly, Heckman and Robb (1986) argue that the econometric approach based on latent index models uses the propensity score "in a different way than that advocated by Rosenbaum and Rubin [1983]." This paper shows that differences between the usage of the propensity score in the econometric and statistics literature are not as great as previously believed. Under mild assumptions, including (but weaker than) those usually invoked in latent index models, the joint distribution of potential outcomes and selection status is independent of exogenous regressors or instruments given the selection propensity score. This result generalizes conditional independence of the Rosenbaum and Rubin (1983) type to selection models. As in Ahn and Powell (1993), this approach leads to estimators that condition on the selection propensity score.

An application of these ideas using quarter of birth to construct instrumental variables estimates of the effect of schooling on earnings illustrates some practical difficulties with non-parametric selection corrections based on the selection propensity score. First, the resulting estimates are likely to be imprecise. Second, there is a range of alternative estimators that exploit the same identifying information. On the plus side, in the application studied here, various approaches to the selection problem generate similar results. This finding supports the conclusion that conventional 2SLS estimates of the returns to schooling using quarter of birth as an instrument to estimate wage equations are not biased by conditioning on employment status.

Table 1: OLS and 2SLS Estimates of the Returns to Schooling

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
dependent variable	educ	wage	wage	wage	wage	educ	wage>0	wage>0	wage>0	wage>0
estimate type	ols rf	ols rf	ols	2sls	2sls	ols rf	ols rf	ols	2sls	2sls
instruments				QOB	QOB*YOB				QOB	QOB*YOB
coefficient										
educ			.072 (.0003)	.110 (.018)	.088 (.015)			.013 (.0002)	.045 (.008)	.040 (.007)
quarter I	-.156 (.015)	-.015 (.003)				-.170 (.144)	-.007 (.0015)			
quarter II	-.106 (.015)	-.011 (.003)				-.115 (.015)	-.005 (.0015)			
quarter III	-.037 (.015)	.002 (.003)				-.036 (.014)	-.0007 (.0015)			
2sls χ^2(dof)				4.5(2)	29.3(29)				.41(2)	18.1(29)
N			391,524					462,311		

Notes: Each equation includes year of birth dummies. Standard errors in parentheses. Quarter-of-birth effects are relative to the fourth quarter.

Table 2: 2sls Estimates Controlling for the Probability of Selection

	full sample					By quartile of $P[w_i=1 \mid Z_i]$				average
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
instruments	QOB	QOB	QOB*YOB	QOB*YOB	QOB*YOB	QOB*YOB	QOB*YOB	QOB*YOB	QOB*YOB	
coefficient										
educ	.110	-.160	.088	.085	.106	-.071	.122	.167	.130	.087
	(.018)	(.192)	(.015)	(.024)	(.026)	(.207)	(.055)	(.073)	(.083)	(.060)
$P[w_i=1 \mid Z_i]$		5.43		.069	-.416					
		(3.83)		(.440)	(.200)					
$P[w_i=1 \mid Z_i]^2 \times 100$					.245					
					(.117)					
Quartiles of $P[w_i=1 \mid Z_i]$						-.0035 → .0	0. → .003	.003 → .008	.008 → .014	
2sls χ^2(dof)	4.46(2)	.01(2)	29.3(29)	29.3(28)	24.3(27)	.09(2)	2.9(29)	.20(29)	1.0(29)	

Notes: The sample includes men with positive earnings only. Each equation includes year of birth dummies. Standard errors in parentheses. Estimates in column (5) are from models that include year dummies when computing the selection propensity score. Estimates in columns (6-9) are computed after stratifying on within YOB variation in the selection propensity score.

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., and G.W. Imbens, (1995), "Two-Stage Least Squares Estimates of Average Causal Response in Models with Variable Treatment Intensity," *Journal of the American Statistical Association* 90, June.

Angrist, J.D., G.W. Imbens, and D. Rubin (1995), "Identification of Causal Effects Using Instrumental Variables," forthcoming, *Journal of the American Statistical Association*.

Angrist, J.D., and A.B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics*, September.

Angrist, J.D., and A.B. Krueger (1992): "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples" *Journal of the American Statistical Association* 87, June.

Ashenfelter, O.A., and J. Ham (1979), "Education, Unemployment, and Earnings," *Journal of Political Economy* 87(5, part 2), S99-S116.

Card, D.E. (1994), "Earnings, Schooling, and Ability Revisited," Industrial Relations Section, Princeton University, Working Paper No. 331, May.

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.

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

Haavelmo, Trygve, 1944, *The Probability Approach in Econometrics*, Supplement to *Econometrica* 12, July 1944, Chicago: The Econometric Society.

Ham, John, and R.J. Lalonde (1995), "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data," forthcoming, *Econometrica*.

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 M.R. Killingsworth (1986), "Female Labor Supply," Chapter 2 in O. Ashenfelter and R. Layard, eds., *The Handbook of Labor Economics Volume I*, Amsterdam: North-Holland.

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.

Imbens, G. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, March.

Kleinbaum, D.G., L.L. Kupper, and H. Morgenstern (1982), *Epidemiological Research*, New York: Van Nostrand Reinhold.

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.

Manski, (1994), "The Selection Problem," Chapter 4 in C. Sims, ed., *Advances in Econometrics Volume I*, 6th World Congress of the Econometric Society, Cambridge: Cambridge University Press.

Mincer, Jacob (1991), "Education and Unemployment," National Bureau of Economic Research Working Paper No. 3838, September.

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., 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.