TECHNICAL WORKING PAPER SERIES

TRIMMING FOR BOUNDS ON TREATMENT EFFECTS WITH MISSING OUTCOMES

David S. Lee

Technical Working Paper 277
http://www.nber.org/papers/T0277

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2002

I thank Jim Powell, David Card, Guido Imbens, Jack Porter, Josh Angrist, Ed Vytlacil, Aviv Nevo, Jonah Gelbach, Doug Miller, David Autor, and participants of the UC Berkeley Econometrics and Labor Lunches, for helpful discussions and suggestions. The views expressed in this paper are those of the author and not necessarily those of the National Bureau of Economic Research.

Trimming for Bounds on Treatment Effects with Missing Outcomes
David S. Lee
NBER Technical Working Paper No. 277
June 2002
JEL No. C10, C24, J20, J00

ABSTRACT

Empirical researchers routinely encounter sample selection bias whereby 1) the regressor of interest is assumed to be exogenous, 2) the dependent variable is missing in a potentially non-random manner, 3) the dependent variable is characterized by an unbounded (or very large) support, and 4) it is unknown which variables directly affect sample selection but not the outcome. This paper proposes a simple and intuitive bounding procedure that can be used in this context. The proposed trimming procedure yields the tightest bounds on average treatment effects consistent with the observed data. The key assumption is a monotonicity restriction on how the assignment to treatment effects selection -- a restriction that is implicitly assumed in standard formulations of the sample selection problem.

David S. Lee
Department of Economics
University of California, Berkeley
549 Evans Hall, #3880
Berkeley, CA 94720-3880
and NBER
dslee@econ.berkeley.edu
510-642-4628

1 Introduction

It is well-known that the econometric identification of causal parameters of interest becomes even more challenging when outcome data are unobserved in a non-random way. In some cases, outcome data is "missing" due to non-response or sample attrition. In other cases, outcomes may not even be well-defined for the entire population. For example, hourly or weekly wages are not defined for the non-working (Heckman, 1974). When the process determining observability of the outcome is related to determinants of the outcome, an analysis that ignores the sample selection process will in general yield biased estimates of the effects of the exogenous regressor of interest (Heckman, 1979). Even the most well-designed randomized experiment or the most compelling quasi-experiment is susceptible to selection bias due to missing outcomes.

There are two general approaches to addressing the problem. One is to explicitly model the process determining selection. In some cases, it involves assuming that data are missing at random, perhaps conditional on a set of covariates (Rubin 1976). Alternatively, it involves assuming the existence of exogenous variables that determine selection, but do not have its own direct impact on the outcome of interest. Such an exclusion restriction is often utilized in parametric and semi-parametric models of the censored selection process (Heckman 1979, 1990; Ahn and Powell 1993; Andrews and Schafgans 1998; Das, Newey, and Vella 2000).

Researchers' reluctance to rely upon specific exclusion restrictions motivates an alternative approach. This approach utilizes boundedness of the support of the outcome variable in order to construct "worst-case" bounds for the treatment effect parameter – bounds that are still consistent with the data that are observed. Horowitz and Manski (2000a) use this notion to provide a general framework for constructing bounds for treatment effect parameters when outcome and covariate data are non-randomly missing in an experimental setting. Others (Balke and Pearl 1997; Heckman and Vytlacil 1999, 2000a, 2000b) have constructed such bounds to address a different problem – that of imperfect compliance of the treatment, even when "intention" to treat is effectively randomized (Bloom 1984; Robins 1989; Imbens and Angrist, 1994; Angrist, Imbens, and Rubin 1996). A limitation of these kinds of procedures is that when outcomes are unbounded (or have very large support), finite (or reasonably informative) bounds for means cannot be generated without some further restriction on the sample selection process (Manski 1995).

This paper proposes a procedure for bounding average treatment effects in the presence of non-randomly missing outcomes, without relying on exclusion restrictions, even when the support of the outcome variable is unbounded. A monotonicity restriction on the sample selection process allows one to "trim" observed distributions of data in order to yield sharp bounds on average treatment effects. The two key assumptions which justify the procedure are 1) "as good as" random assignment of treatment (independence between the regressor of interest and the errors in the outcome and selection equations) and 2) a monotonicity condition – whereby assignment to treatment impacts selection probabilities only in "one direction". The first assumption is commonly adopted by both the existing modeling and bounding approaches, and the second is also implicitly assumed in existing approaches that explicitly model the sample selection process. The procedure can be directly applied, for example, to the analysis of randomized experiments in which there is missing outcome data.

In addition, the discussion below makes it clear that, given unbounded outcomes, these two assumptions are *not* sufficient for generating bounds on the average treatment effect for the *entire* population. Bounds can only be generated for a specific sub-population: individuals whose outcomes will be observed, irrespective of the assignment to treatment. As shown below, in some contexts, this effect may in fact be the parameter of interest. However, when bounds of the effects for other sub-populations (e.g. those whose outcomes will *not* be observed, irrespective of the assignment to treatment) are the objects of interest, *further* restrictions on the sample selection process are necessary.

The paper is organized as follows. Section 2 describes the basic model and trimming procedure, providing economic examples in which the above average treatment effect is the parameter of interest, and in which the monotonicity condition will or will not hold. Section 3 describes how baseline covariates can be used narrow the width of the bounds. Section 4 discusses some testable implications of the key restrictions of the model for trimming, and Section 5 concludes. Throughout this paper, the treatment variable is assumed to be dichotomous, and always observed; hence, the analysis applies to censored and not truncated samples.

2 Missing Outcomes in a Heterogeneous Treatment Effect Model

I begin by outlining conditions under which a trimming approach can produce bounds for average treatment effects for a specific sub-population of interest. Consider the random variables $(Y_1^*, Y_0^*, S_1, S_0, D)$ where Y_1^* and Y_0^* are continuous and unbounded potential outcomes of interest when $D = 1$ and $D = 0$, respectively. S_1 and S_0 denote whether the outcome is observed when $D = 1$ and $D = 0$, respectively. For example, the realization $S_1 = 1, S_0 = 0$ implies that the outcome would be observed if $D = 1$, but would be missing if $D = 0$. (Y, S, D) is observed, where $Y = Y_1^* D + Y_0^* (1 - D)$ if $S = 1$, Y is missing if $S = 0$; also, $S = S_1 D + S_0 (1 - D)$. Y_1^* and Y_0^* are never simultaneously observed, and S_1 and S_0 are never simultaneously observed.

Assumption A

$$(Y_1^*, Y_0^*, S_1, S_0) \text{ is independent of } D \tag{1}$$

This assumption corresponds to the “as good as” random assignment of D. It is useful to consider this assumption, as it means that any bias in identifying average treatment effects will be due to censored selection, rather than to the usual confounding problem.

Furthermore, it is assumed that assignment to D, if it affects S at all, can affect S in only “one direction”. This is a “monotonicity” assumption.

Assumption B

$$\Pr [S_1 = 0, S_0 = 1] = 0 \tag{2}$$

This assumption precludes the possibility that within a population of interest, some individuals are induced to drop out of the sample because of the treatment. It is important to note that the choice of imposing $\Pr [S_1 = 0, S_0 = 1] = 0$ rather than $\Pr [S_1 = 1, S_0 = 0] = 0$ is innocuous. I consider this case for expositional purposes, and a parallel argument to that presented below is valid if the latter assumption is imposed instead. This assumption is analogous to the monotonicity assumption in studies of imperfect compliance of treatment (Imbens and Angrist 1994; Angrist, Imbens, and Rubin 1996).

Assumptions A and B imply that the difference between the means of the sample-selected treatment and control groups is

$$\begin{aligned} & E\left[Y|D=1,S=1\right]-E\left[Y|D=0,S=1\right] \\ = \; & \frac{\Pr\left[S_0=0,S_1=1|D=1\right]}{\Pr\left[S=1|D=1\right]}E\left[Y_1^*|S_0=0,S_1=1\right] \\ & +\frac{\Pr\left[S_0=1,S_1=1|D=1\right]}{\Pr\left[S=1|D=1\right]}E\left[Y_1^*|S_0=1,S_1=1\right] \\ & -E\left[Y_0^*|S_0=1,S_1=1\right] \end{aligned} \qquad (3)$$

In general, this will be biased for a particular parameter of interest: $E\left[Y_1^*-Y_0^*|S_0=1,S_1=1\right]=E[Y_1^*|S_0=1,S_1=1]-E[Y_0^*|S_0=1,S_1=1]$, the average treatment effect for the subpopulation whose outcome data will be observed irrespective of treatment status. While the weights $\frac{\Pr[S_0=0,S_1=1|D=1]}{\Pr[S=1|D=1]}$ and $\frac{\Pr[S_0=1,S_1=1|D=1]}{\Pr[S=1|D=1]}$ can be identified from the observed data, $E\left[Y_1^*|S_0=0,S_1=1\right]$ and $E[Y_1^*|S_0=1, S_1=1]$ cannot be identified without further restrictions.

However, without further restrictions, the observed data can yield upper and lower *bounds* $\overline{E}$ and $\underline{E}$ such that $\underline{E}\leq E[Y_1^*|S_0=1,S_1=1]\leq\overline{E}$. It follows that there exist bounds such that

$$\underline{E}-E\left[Y|D=0,S=1\right]\leq E\left[Y_1^*-Y_0^*|S_0=1,S_1=1\right]\leq\overline{E}-E\left[Y|D=0,S=1\right] \qquad (4)$$

for the average treatment effect for this subpopulation.

The approach in this paper is to construct these bounds by trimming the lower or upper tails of the observed distribution of Y for the treatment group, by a proportion given by $\frac{\Pr[S=1|D=1]-\Pr[S=1|D=0]}{\Pr[S=1|D=1]}$: the proportion of the selected treatment group that is induced to have a non-missing value of the outcome because of the assignment to treatment.

Proposition 1 *Suppose Assumptions A and B hold, and* $\Pr\left[S=1|D=0\right]\neq 0$. *Denote the observed density and cumulative distribution of* Y, *conditional on* $D=1$ (*and* $S=1$), *as* $f\left(y\right)$ *and* $F\left(y\right)$, *respectively. Then*

$$\underline{E}\equiv\frac{1}{1-p}\int_{-\infty}^{F^{-1}(1-p)}yf\left(y\right)dy\leq E\left[Y_1^*|S_0=1,S_1=1\right]$$

and

$$\overline{E}\equiv\frac{1}{1-p}\int_{F^{-1}(p)}^{\infty}yf\left(y\right)dy\geq E\left[Y_1^*|S_0=1,S_1=1\right]$$

where

$$p = \frac{\Pr[S=1|D=1] - \Pr[S=1|D=0]}{\Pr[S=1|D=1]}$$

Also, $\underline{E}$ $\left(\overline{E}\right)$ *is equal to the smallest (largest) possible value for* $E[Y_1^*|S_0=1, S_1=1]$ *that is consistent with the distribution of observed data on* (Y, S, D)*.*

Given Assumption B, $E[Y_0^*|S_0=1, S_1=1]$ equals $E[Y|D=0, S=1]$, which can be computed from the observed data from the control group.

Corollary 2 *Given Assumptions A and B and* $\Pr[S=1|D=0] \neq 0$

$$\underline{E} - E[Y|D=0, S=1] \leq E[Y_1^* - Y_0^*|S_0=1, S_1=1] \leq \overline{E} - E[Y|D=0, S=1]$$

where the lower bound (upper bound) is the smallest (largest) possible value for the average treatment effect, $E[Y_1^* - Y_0^*|S_0=1, S_1=1]$*, that is consistent with the distribution of the observed data on* (Y, S, D)*.*

The "monotonicity" assumption is crucial to this approach. It ensures that subpopulation of the control group for whom we observe outcomes consists only of those for whom $S_0 = 1, S_1 = 1$ – that is, those who will always have non-missing outcome data, irrespective of the assignment to treatment. Without monotonicity, the control and treatment groups could consist solely of the sub-populations for whom $S_0 = 1, S_1 = 0$ and $S_0 = 0, S_1 = 1$, respectively. This would imply no "overlap" between the two sub-populations, making it impossible to make a comparison that could be interpreted as a causal effect. The independence assumption is also important, since it is what justifies the contrast between the trimmed population of the treatment group and the control group.

The following are two economic examples of when the parameter $E[Y_1^* - Y_0^*|S_0=1, S_1=1]$ is of economic interest. In the first, the monotonicity condition could be expected to hold, and in the second, economic reasoning suggests that monotonicity would probably not hold.

Example 1 *Labor supply with a Negative Income Tax, experimental variation in tax rate*

Consider a static labor supply setting, where we are interested in the *intensive* margin response of hours of work to a change in marginal tax rates. Subjects are randomized into treatment and control groups. Both groups are given the same guaranteed income subsidy of G, which is taxed away at rates t_t and $t_c(> t_t)$, for the treated and control, respectively. Suppose we are interested in the average treatment effect of the experimental variation in the tax rate on Y, the natural logarithm of hours worked. Obviously,

Y will be undefined for the nonworking, and we might expect the treatment (a higher effective wage) to induce some individuals to work, causing a potential sample selection bias.

Under the assumption of optimizing behavior given a complete, transitive, and strictly monotone preference relation over leisure and consumption (l and c), any consumer who would work positive hours facing tax rate t_c would work positive hours facing t_t. To see this, consider any individual who works positive hours under t_c. Denote the optimal hours as $\exp\left(Y_0^*\right) = h_0 > 0$. The bundle of consumption and leisure $\left(G + w\left(1 - t_t\right) h_0, T - h_0\right)$, (where T is total time available), which is a bundle that is feasible given the treatment, is strictly preferred to $\left(G + w\left(1 - t_c\right) h_0, T - h_0\right)$, which itself is preferred to (G, T) (the bundle attained by not working) by hypothesis. By transitivity, (G, T) cannot be the optimal choice for the consumer facing t_t.

Thus, in this economic context, the monotonicity assumption (**B**) is rationalized by optimizing behavior given a fairly standard preference relation. The trimming procedure described above can be used to generate bounds on the percentage change in hours of labor supply induced by a marginal tax rate reduction, accounting for the presence of non-random sample selection that results from labor supply behavior on the extensive margin of employment.

Example 2 *Labor supply with a Negative Income Tax, experimental variation in tax rate* and *guaranteed subsidy*

Consider the same setting as above, except that in addition to different tax rates, different levels of the guaranteed subsidy $G_t > G_c$ are offered to the treatment and control groups, respectively. Again, consider the control group individual who optimally chooses positive hours of work by choosing the combination $\left(G_c + w\left(1 - t_c\right) h_0, T - h_0\right)$. Without further information about preferences, we cannot rule out the possibility that (G_t, T) is strictly preferred by this individual, and that it would have been the optimal choice under the treatment assignment. In other words, we cannot rule out the possibility that treatment *induces* some individuals to stop working. We also cannot rule out that the treatment induces other individuals to work positive hours (in other words, that $\left(G_t + w\left(1 - t_t\right) h_1, T - h_1\right)$, $h_1 > 0$, is preferred to (G_t, T) (which, in turn, is preferred to (G_c, T)). In this example, economic reasoning *cannot* be used to justify Assumption **B**.

It should also be noted that the independence and monotonicity conditions are implicitly assumed within typical latent-variable formulations of the sample selection process (as in Heckman 1979). Consider the system of equations

$$\begin{aligned} Y^* &= \beta_0 + \beta_1 T + U \\ Z^* &= \gamma_0 + \gamma_1 T + V \end{aligned} \tag{5}$$

where Y^* is an outcome of interest, T takes on the values 0 or 1, β_1 is the treatment effect of interest. Y is observed and equals Y^* if $Z^* \geq 0$, but is missing if $Z^* \leq 0$. It is often assumed (for example, in maximum likelihood estimation of parametric selection models) that (U, V) is independent of T. In addition, if $\gamma_1 \geq 0$, then it is possible to use the bounds proposed above to assess missing outcome bias. To see this, note that this system implies $Y_1^* = \beta_0 + \beta_1 + U$, $Y_0^* = \beta_0 + U$, $S_1 = 1\left(V \geq -\gamma_0 - \gamma_1\right)$, $S_0 = 1\left(V \geq -\gamma_0\right)$, where $1\left(A\right)$ is an indicator variable that equals 1 in the event of A (0 otherwise), and $D = T$. The independence of T implies Assumption A, and if $\gamma_1 \geq 0$, then $\Pr\left(V < -\gamma_0 - \gamma_1, V \geq -\gamma_0\right) = 0$, implying that Assumption B holds also. It should be noted that the proposed bounding procedure is applicable to a more general "heterogeneous treatment effect" version of the above latent-variable formulation. This is because the independence and monotonicity assumptions are equivalent to a generalized latent index, threshold-crossing model (Vytlacil 2000).

An important implication of Assumptions A and B is that as p vanishes, so does the sample selection bias. The intuition is that if $p = 0$, then under the monotonicity assumption, the population with observed outcome data – whether in the treatment or control group – is comprised of individuals whose *sample selection* was unaffected by the assignment to treatment (those for whom $S_0 = 1$, and $S_1 = 1$). These individuals can be thought of as the "always-takers" sub-population (Angrist, Imbens, and Rubin 1996), except that "taking" is not the taking of the treatment, but rather selection into the sample. One example of a practical implication of this is that when analyzing randomized experiments, if the "drop-out" rates in the treatment and control groups are similar, and if the monotonicity condition is believed to hold, then a comparison of the treatment and control means is a valid estimate of an average treatment effect.

The notion that there is no sample selection bias when the probability of selection is the same in the treated and control groups can also be seen by examining the dichotomous treatment case within the frameworks that condition on an unknown function of the probability of selection (as in Heckman and Robb 1986; Heckman 1990; Ahn and Powell 1993; Angrist 1997; Andrews and Schafgans 1998; Das, Newey, and Vella 2000). However, in these studies, it is clear that when the treated and control group selection probabilities are different, point identification is lost without imposing an exclusion restriction on auxiliary variables that determine selection. Thus, the bounding or "sensitivity" analysis proposed here can be viewed as an alternative to hypothesizing the existence of such auxiliary variables, that are needed to achieve point identification.

It is instructive to highlight the primary features of the proposed trimming procedure that distinguish it from existing bounds approaches in the literature. First, the model and procedure proposed here can produce finite bounds when the outcome has unbounded support. This should be contrasted to a method that addresses missing outcomes by essentially assigning the values of upper and lower bounds of support to missing data to bound parameters of interest (Horowitz and Manski 1998, 2000a).

This advantage of trimming, however, does not come without a cost. The second distinctive feature (and disadvantage) of the model proposed above is that it relies crucially on an unverifiable assumption about the selection process. For example, the model assumes that *every* control (treatment) group individual who reported an outcome would have reported outcome if they had been assigned treatment (to the control group) – a conjecture that simply cannot be verified one way or another. The appropriateness of this "monotonicity" assumption may or may not be "plausible" depending on the particular application, as illustrated in the economic examples above.

A third distinctive feature is that the bounds can only be generated for the average treatment effects for a specific sub-population: those individuals whose outcomes will be observed, irrespective of the assignment to treatment. Sometimes, that parameter may be of interest (as in the economic examples described above), but in other situations, one may be interested in average treatment effects for the other two sub-populations: 1) those that were "induced" to yield valid outcome data because of the treatment, $E\left[Y_1^* - Y_0^* | S_0 = 0, S_1 = 1\right]$, and 2) those that will always have missing outcomes, irrespective of the treatment status, $E\left[Y_1^* - Y_0^* | S_0 = 0, S_1 = 0\right]$. In those cases, it is clear that the independence and monotonicity conditions will *not* be sufficient for generating informative bounds for those effects. Further stochastic restrictions would be necessary.

3 Trimming Using Baseline Covariates

Researchers often possess "baseline" characteristics of both the treatment and control subjects. When analyzing randomized experiments, these covariates are typically used to assess whether or not the randomization "failed", and if successful randomization is not rejected by the data the covariates are often included in the analysis to reduce the sampling variability of the estimates. These covariates can be used in a modified trimming method that will lead to tighter bounds on $E\left[Y_1^* - Y_0^* | S_0 = 1, S_1 = 1\right]$ than that constructed without the covariates. I suppose that there is no missing data on these baseline covariates, in contrast to the generalized bounds analysis of Horowitz and Manski (2000a).

Suppose there exists a vector of baseline covariates X, where each element has discrete support, so that this vector can take on one of a finite number of discrete values. Focus on the values $\{x_1, \ldots, x_J\}$, such that for each $j = 1, \ldots, J$, $\Pr\left(X = x_j | D = 0, S = 1\right) \neq 0$.

Assumption C

$$\left(Y_1^*, Y_0^*, S_1, S_0, X\right) \text{ is independent of } D \qquad (6)$$

Assumption C would hold if D were randomly assigned, and X were pre-determined, relative to the point of random assignment.

Under this assumption, an upper (lower) bound for $E\left[Y_1^* - Y_0^* | S_0 = 1, S_1 = 1\right]$ can also be constructed by trimming the lower (upper) tails of distributions of y, conditional on $D = 1$ and X, by a proportion given by $p_j = \frac{\Pr[S=1|D=1,X=x_j]-\Pr[S=1|D=0,X=x_j]}{\Pr[S=1|D=1,X=x_j]}$. The overall mean of the truncated distributions of the sub-groups of the treated is computed by averaging across values of X.

Proposition 3 *Suppose Assumptions B and C hold, and* $\Pr\left[S = 1 | D = 0\right] \neq 0$. *Denote the observed* *density and cumulative distribution of* Y*, conditional on* $D = 1$ *(and* $S = 1$*) and* $X = x_j$*, as* $f(y|x_j)$ *and* $F(y|x_j)$*, respectively. Then*

$$\underline{E}^* \equiv \sum_{j=1}^{J} \Pr[X = x_j | S = 1, D = 0] \frac{1}{1-p_j} \int_{-\infty}^{F^{-1}(1-p_j|x_j)} y f(y|x_j)\, dy \leq E[Y_1^* | S_0 = 1, S_1 = 1]$$

and

$$\overline{E}^* \equiv \sum_{j=1}^{J} \Pr[X = x_j | S = 1, D = 0] \frac{1}{1-p_j} \int_{F^{-1}(p_j|x_j)}^{\infty} y f(y|x_j)\, dy \geq E[Y_1^* | S_0 = 1, S_1 = 1]$$

where

$$p_j = \frac{\Pr[S = 1 | D = 1, X = x_j] - \Pr[S = 1 | D = 0, X = x_j]}{\Pr[S = 1 | D = 1, X = x_j]}$$

Also, $\underline{E}^*$ $\left(\overline{E}^*\right)$ *is equal to the smallest (largest) possible value for* $E[Y_1^* | S_0 = 1, S_1 = 1]$ *that is consistent with the distribution of observed data on* (Y, S, D, X)

Corollary 4 *Given Assumptions B and C and* $\Pr[S = 1 | D = 0] \neq 0$

$$\underline{E}^* - E[Y | D = 0, S = 1] \leq E[Y_1^* - Y_0^* | S_0 = 1, S_1 = 1] \leq \overline{E}^* - E[Y | D = 0, S = 1]$$

where the lower bound (upper bound) is the smallest (largest) possible value for the average treatment effect, $E[Y_1^* - Y_0^* | S_0 = 1, S_1 = 1]$*, that is consistent with the distribution of the observed data on* (Y, S, D, X)*.*

Intuitively, Assumption C implies that the assumptions used to justify the trimming procedure will also justify trimming, conditional on X. Given bounds for $E[Y_1^* - Y_0^* | S_0 = 1, S_1 = 1, X = x_j]$, it is possible to average across values of X to produce bounds for $E[Y_1^* - Y_0^* | S_0 = 1, S_1 = 1]$.

The motivation for this modified trimming procedure is that using the covariates in this way will lead to tighter bounds on the treatment effect parameter of interest.

Proposition 5 *If Assumptions B and C hold and* $\Pr[S = 1 | D = 0] \neq 0$ *, then* $\underline{E}^* \geq \underline{E}$ *and* $\overline{E}^* \leq \overline{E}$*.*

Intuitively, this is true because a lower-tail truncated mean of a distribution will always be larger than the average of lower-tail truncated means of sub-groups of the population, provided that the proportion of the entire population that is eventually truncated remains fixed. An implication of Proposition 5 is that in general, using more baseline covariates will lead to producing tighter bounds on $E[Y_1^* - Y_0^* | S_0 = 1, S_1 = 1]$.

It is interesting to relate these trimming bounds to the estimand that would result from a "matching on observables" approach to addressing missing outcome bias. Matching on the baseline covariates would dictate computing the quantity $\sum_{j=1}^{J} \Pr[X = x_j | S = 1, D = 0] \{E[Y | D = 1, S = 1, X = x_j] -$

$E\left[Y|D=0,S=1,X=x_j\right]\}$. A comparison with the comparable quantity in the Corollary above makes it clear that this quantity will lie strictly in between the upper and lower "trimming" bounds.

4 Testable Implications

While it is clear that the assumptions of the model proposed above are fundamentally unverifiable, it is important to examine whether the restrictions generate any testable implications, however weak they might be. As is well known, the independence assumption (C), which corresponds to random assignment, has the implication that the baseline pre-determined characteristics X be distributed identically between the treatment and control groups.

The monotonicity assumption (B) is restrictive enough to generate a testable restriction. In particular, Assumption **B** implies that there exists no j, such that $\Pr\left[S=1|D=1,X=x_j\right] < \Pr[S = 1|D = 0,\ X = x_j]$. Essentially, the monotonicity restriction is inconsistent with the existence of j' and j'' such that $\Pr\left[S=1|D=1,X=x_{j'}\right] < \Pr[S = 1|D = 0,\ X = x_{j'}]$ while at the same time $\Pr\left[S=1|D=1,X=x_{j''}\right] > \Pr\left[S=1|D=0,X=x_{j''}\right]$.

Finally, suppose $\Pr\left[S=1|D=1\right] = \Pr\left[S=1|D=0\right]$. As mentioned earlier, in this case, Assumptions B and C imply that there is no sample selection bias, and that a simple contrast between $E\left[Y|D=1,S=1\right] - E\left[Y|D=0,S=1\right]$ is valid for identifying a meaningful causal parameter. $0 = \Pr\left[S=1|D=1\right] - \Pr\left[S=1|D=0\right] = \sum_j^J\{\Pr\left[X=x_j|D=0\right]\,(\Pr[S=1|D=1,X=x_j] - \Pr[S = 1|\ D=0,X=x_j])\}$ because of Assumption C. Since $\Pr\left[X=x_j|D=0\right] > 0$ for all $j = 1,\ldots,J$, and Assumption B implies that $\Pr\left[S=1|D=1,X=x_j\right] - \Pr\left[S=1|D=0,X=x_j\right] \geq 0$ for $j=1,\ldots,J$, then it must be true that $\Pr\left[S=1|D=1,X=x_j\right] - \Pr\left[S=1|D=0,X=x_j\right] = 0$ for $j=1,\ldots,J$. It can then be shown, using Assumption C and Bayes' Rule, that this implies $\Pr\left[X=x_j|S=1,D=1\right] = \Pr\left[X=x_j|S=1,D=0\right]$ for $j = 1,\ldots,J$. Therefore, if $\Pr\left[S=1|D=1\right] = \Pr\left[S=1|D=0\right]$, then Assumptions B and C imply that the distributions of the baseline covariates between the selected treatment group and the selected control group are identical, which is testable given the observed data.

5 Conclusions and Extensions

In many situations, researchers may be willing to entertain the possibility that treatments are "as good as randomly assigned" but are at the same time considerably less confident about the underlying process that determines whether outcomes are missing. A potentially useful alternative to specifying exclusion restrictions is a bounding analysis that takes generates "worst-case" sample selection biases. In the context of outcomes with essentially unbounded support, existing nonparametric bounding approaches (e.g. Horowitz and Manski 1998, 2000a) of unbounded outcomes immediately suggest there will be no finite bounds on average treatment effects. This can be informative in the sense that it suggests that *any* finite bounds on treatment effects in this context will necessarily be a consequence of some further stochastic restriction on the data generating process (Horowitz and Manski 2000b). The question then becomes Which restrictions have relatively large benefits and/or small costs?

This paper has proposed a simple and intuitive trimming procedure that is justified under the added restriction of monotonicity of the censored selection process. The main benefit from imposing this restriction is that it allows one to generate finite bounds even when the outcome variable has unbounded support. The main cost of the restriction is that such a behavioral assumption may or may not be plausible, depending on the particular context of the selection problem. This paper has described two economic contexts: one in which the monotonicity assumption could be considered plausible, and another where economic reasoning suggests the assumption is unwarranted.

The following are potentially useful avenues for future research. First, it would be interesting to apply the proposed trimming procedure to appropriate applied contexts, and to compare the bounds to estimates obtained from other parametric and semi-parametric modeling approaches and other bounding procedures. Second, since the number of baseline covariates may be so large as to create a "small cell" problem, it would be helpful to generalize the procedure to utilize continuous covariates. Third, it seems possible to generalize the procedure in various directions. For example, it could be extended to apply to 1) the case of an endogenous regressor of interest with a valid instrument (or imperfect compliance of a treatment whose “intention-to-treat” is randomized), 2) the case of a continuous treatment variable, or 3) the case of more than one sample selection process (e.g. sample attrition as distinct from the labor force participation decision). Finally, it would be interesting to explore what additional plausible assumptions, beyond the monontonicity restriction, would lead to tighter bounds on average treatment effects.

Appendix A.

Lemma 6 *Suppose the probability density $f^*(y)$ is a mixture of two probability densities, $m^*(y)$ and $n^*(y)$ such that $f^*(y) = p^*m^*(y) + (1-p^*)\, n^*(y)$, where $p^* \in [0,1)$ is fixed. Let $F^*(y)$ be the cumulative distribution function corresponding to $f^*(y)$. Consider the truncated density $g^*(y)$ which is equal to $\frac{1}{1-p^*} f^*(y)$ on $\left[F^{*-1}(p^*), \infty\right]$, 0 otherwise. Then $\int_{-\infty}^{\infty} yg^*(y)\, dy \geq \int_{-\infty}^{\infty} yn^*(y)\, dy$.*

Proof of Lemma 6. First consider $p^* \in (0,1)$. Let $N^*(y)$ be the cumulative distribution function corresponding to $n^*(y)$. First, compare the truncated density, $g^*(y)$ to an arbitrarily chosen $n^*(y)$ that is not identical to $g^*(y)$. $\int_{-\infty}^{\infty} yg^*(y)\, dy - \int_{-\infty}^{\infty} yn^*(y)\, dy = \int_{F^{*-1}(p^*)}^{\infty} y \left(\frac{1}{1-p^*}\right) f^*(y)\, dy - \int_{-\infty}^{\infty} yn^*(y)\, dy = \int_{F^{*-1}(p)}^{\infty} y \left[\left(\frac{1}{1-p^*}\right) f^*(y) - n^*(y)\right] dy - \int_{-\infty}^{F^{*-1}(p^*)} y \cdot\ n^*(y)\, dy$. Multiplying both sides by $\frac{1}{N^*(F^{*-1}(p^*))}$ yields $\frac{1}{N^*(F^{*-1}(p^*))} \left\{\int_{-\infty}^{\infty} yg^*(y)\, dy - \int_{-\infty}^{\infty} yn^*(y)\, dy\right\} = \frac{1}{N^*(F^{-1}(p^*))} \int_{F^{*-1}(p^*)}^{\infty} y[\left(\frac{1}{1-p^*}\right) f^*(y) - n^*(y)]dy - \frac{1}{N^*(F^{*-1}(p^*))} \int_{-\infty}^{F^{*-1}(p)} yn^*(y)\, dy$. By definition $n^*(y) = \frac{f^*(y) - p^*m^*(y)}{1-p^*}$, so for any y on $\left[F^{*-1}(p^*), \infty\right]$, $\frac{1}{1-p} f^*(y) - n^*(y) \geq 0$. If $n^*(y) \neq g^*(y)$, then it can be shown that $\frac{1}{N^*(F^{*-1}(p^*))}$ $\left[\left(\frac{1}{1-p^*}\right) f^*(y) - n^*(y)\right]$ defined on $\left[F^{*-1}(p^*), \infty\right]$ and $\frac{1}{N^*(F^{-1}(p))} n^*(y)$ defined on $\left[-\infty, F^{*-1}(p^*)\right]$ are each proper probability densities that integrate to 1. The support of the former is strictly above the support of the latter. Therefore, $\frac{1}{N^*(F^{*-1}(p^*))} \left\{\int_{-\infty}^{\infty} yg^*(y)\, dy - \int_{-\infty}^{\infty} yn^*(y)\, dy\right\} > 0$. Second, consider the case that $n^*(y) = g^*(y)$. Then $\int_{-\infty}^{\infty} yg^*(y)\, dy - \int_{-\infty}^{\infty} yn^*(y)\, dy = 0$.

Now consider $p^* = 0$. Then $g^*(y) = f^*(y) = n^*(y)$, so $\int_{-\infty}^{\infty} yg^*(y)\, dy = \int_{-\infty}^{\infty} yn^*(y)\, dy$.

Proof of Proposition 1. Assumption A and B implies that $p = \frac{\Pr[S=1|D=1] - \Pr[S=1|D=0]}{\Pr[S=1|D=1]} = \frac{\Pr[S_0=0, S_1=1|D=1]}{\Pr[S=1|D=1]}$. p is strictly less than 1 by assumption. Assumption B also implies that $f(y) = pm(y) + (1-p)\, n(y)$, where $m(y)$ denotes the density of Y_1^*, conditional on $D = 1$, $S_0 = 0$, $S_1 = 1$, and $n(y)$ denotes the density of Y_1^*, conditional on $D = 1$, $S_0 = 1$, $S_1 = 1$. By Assumption A, $n(y)$ is also the density of Y_1^*, conditional on $S_0 = 1, S_1 = 1$. By Lemma 6, $\overline{E} \equiv \frac{1}{1-p} \int_{F^{-1}(p)}^{\infty} yf(y)\, dy \geq \int_{-\infty}^{\infty} yn(y)\, dy = E\left[Y_1^*|S_0 = 1, S_1 = 1\right]$.

To show that $\overline{E}$ equals the maximum possible value for $E\left[Y_1^*|S_0 = 1, S_1 = 1\right]$ that is consistent with the distribution of the observed data on (Y, S, D), note first that the observed data can be completely described by 1) $f(y)$, 2) the density of Y conditional on $S = 1, D = 0$, and 3) the probability function $\Pr[S = s, D = d]$, $s, d = 0, 1$. By Assumptions A and B, the density of Y conditional on $S = 1, D = 0$

is equal to the density of Y_0^* conditional on $S_0 = 1, S_1 = 1$. Set $n(y)$ equal to the density $\frac{1}{1-p} f(y)$ defined on $\left[F^{-1}(p), \infty\right]$, and $m(y)$ equal to the density $\frac{1}{p} f(y)$ defined on $\left[-\infty, F^{-1}(p)\right]$ where $p \equiv \frac{\Pr[S=1|D=1]-\Pr[S=1|D=0]}{\Pr[S=1|D=1]} = 1 - \frac{\left(1+\frac{\Pr[S=0,D=1]}{\Pr[S=1,D=1]}\right)}{\left(1+\frac{\Pr[S=0,D=0]}{\Pr[S=1,D=0]}\right)}$; there is only one p consistent with the probability function $\Pr[S = s, D = d]$, $s, d = 0, 1$. These choices for $n(y)$ and $m(y)$ are consistent with $f(y)$ satisfying $f(y) = pm(y) + (1-p)n(y)$. Then $E[Y_1^*|S_0 = 1, S_1 = 1]$ will equal $\frac{1}{1-p}\int_{F^{-1}(p)}^{\infty} yf(y)\,dy \equiv \overline{E}$, and it has already been shown that $\overline{E} \geq E[Y_1^*|S_0 = 1, S_1 = 1]$.

An argument parallel to that made above can be made for $\underline{E}$.

Proof of Proposition 3. Given Assumption C, this implies that Assumption A holds, conditionally on X. It is given that for each j, $\Pr[X = x_j|D = 0, S = 1] \neq 0$. So $\Pr[S = 1|D = 0] \neq 0$ implies, using Bayes' Rule, that $\Pr[S = 1|D = 0, X = x_j] \neq 0$ for all $j = 1, \ldots, J$. Thus, by the Proposition 1, it can be shown that $\frac{1}{1-p_j}\int_{F^{-1}(p_j|x_j)}^{\infty} yf(y|x_j)\,dy \geq E[Y_1^*|S_0 = 1, S_1 = 1, X = x_j]$ for $j = 1, \ldots, J$. It follows that $\overline{E}^* \geq \sum_{j=1}^{J} \Pr[X = x_j|S = 1, D = 0]\, E[Y_1^*|S_0 = 1, S_1 = 1, X = x_j]$. The latter quantity equals $\sum_{j=1}^{J}\{\Pr[X = x_j|S_0 = 1, S_1 = 1]E[Y_1^*|S_0 = 1, S_1 = 1, X = x_j]\} = E[Y_1^*|S_0 = 1, S_1 = 1]$ by Assumptions B and C.

To show that $\overline{E}^*$ is equal to the largest possible value for $E[Y_1^*|S_0 = 1, S_1 = 1]$ that is consistent with the distribution of observed data on (Y, S, D, X), note first that the data can be completely described by 1) $f(y|x_j)$, $j = 1, \ldots, J$, 2) the densities of Y conditional on $S = 1, D = 0, X = x_j$, 3) the probability function $\Pr[S = s, D = d|X = x_j]$, $s, d = 0, 1$, and 4) the probability function $\Pr[X = x_j]$, $j = 1, \ldots, J$. Since Assumptions A and B hold conditionally on X, by Proposition 1, $\frac{1}{1-p_j}\int_{F^{-1}(p_j|x_j)}^{\infty} yf(y|x_j)\,dy$ is equal to the largest possible value for $E[Y_1^*|S_0 = 1, S_1 = 1, X = x_j]$ consistent with the observed data on (Y, S, D), conditional on $X = x_j$, for each $j = 1, \ldots, J$.

$\Pr[X = x_j|S_0 = 1, S_1 = 1] = \Pr[X = x_j|S = 1, D = 0]$, by assumptions B and C, and $\Pr[X = x_j|S = 1, D = 0]$ is uniquely determined by the probability functions $\Pr[S = s, D = d|X = x_j]$, $s, d = 0, 1$, and $\Pr[X = x_j]$, $j = 1, \ldots, J$ since $\Pr[X = x_j|S = 1, D = 0] = \frac{\Pr[S=1,D=0|X=x_j]\Pr[X=x_j]}{\sum_{k=1}^{J}\Pr[S=1,D=0|X=x_j]\Pr[X=x_k]}$ by Bayes' Rule. Therefore $\overline{E}^*$ is equal to the largest possible value for $\sum_{j=1}^{J} \Pr[X = x_j|S = 1, D = 0]\, E[Y_1^*|S_0 = 1, S_1 = 1, X = x_j] = \sum_{j=1}^{J} \Pr[X = x_j|S_0 = 1, S_1 = 1]\, E[Y_1^*|S_0 = 1, S_1 = 1, X = x_j] = [EY_1^*|S_0 =

$1, S_1 = 1]$ that is consistent with the observed data on (Y, S, D, X).

An argument parallel to that made above can be made for $\underline{E}^*$.

Proof of Proposition 5. As shown in the beginning of the proof of Proposition 3, Assumptions B and C and $\Pr[S = 1|D = 0] \neq 0$ implies that $\Pr[S = 1|D = 0, X = x_j] \neq 0$ for all $j = 1, \ldots, J$. Therefore $p_j \in [0, 1)$ for $j = 1, \ldots, J$. Let $g(y|x_j) = \frac{1}{1-p_j} f(y|x_j)$ on $\left[F^{-1}(p_j|x_j), \infty\right]$, 0 otherwise. Let $h(y|x_j) = 1(p_j > 0) \cdot \frac{1}{p_j} f(y|x_j)$ on $\left[-\infty, F^{-1}(p_j|x_j)\right]$, 0 otherwise. By construction, $f(y) = \sum_{j=1}^J \Pr[X = x_j|S = 1, D = 1] f(y|x_j) = \sum_{j=1}^J \Pr[X = x_j|S = 1, D = 1] p_j h(y|x_j) + \sum_{j=1}^J \Pr[X = x_j|S = 1, D = 1](1 - p_j) g(y|x_j)$. Let $\widehat{p} = \sum_{j=1}^J \Pr[X = x_j|S = 1, D = 1] p_j$; since $p_j \in [0, 1)$ for $j = 1, \ldots, J$, $\widehat{p}$ also lies on $[0, 1)$. Then $f(y)$ can be re-written as $\widehat{p} m^*(y) + (1 - \widehat{p}) n^*(y)$, where $m^*(y) = \frac{1}{\widehat{p}} \sum_{j=1}^J \{\Pr[X = x_j|S = 1, D = 1] \cdot \ p_j h(y|x_j)\}$ and $n^*(y) = \frac{1}{1-\widehat{p}} \sum_{j=1}^J \Pr[X = x_j|S = 1, D = 1] \cdot (1 - p_j) g(y|x_j)$.

Consider first $\widehat{p} \in (0, 1)$. Since $m^*(y)$ and $n^*(y)$ are both probability densities that integrate to 1, Lemma 6 applies: $\frac{1}{1-\widehat{p}} \int_{F^{-1}(\widehat{p})}^{\infty} y f(y)\, dy \geq \int_{-\infty}^{\infty} y n^*(y)\, dy$. All that needs to be shown is that 1) $\widehat{p} = p$, 2) $n^*(y) = \sum_{j=1}^J \Pr[X = x_j|S = 1, D = 0] g(y|x_j)$. If these two statements are true, then $\overline{E} = \frac{1}{1-\widehat{p}} \int_{F^{-1}(\widehat{p})}^{\infty} y f(y)\, dy \geq \int_{-\infty}^{\infty} y n^*(y)\, dy = \overline{E}^*$.

The definition of p_j implies $\widehat{p} = \sum_{j=1}^J \{\Pr[X = x_j|\ S = 1, D = 1](1 - \frac{\Pr[S=1,D=0,X=x_j]\Pr[D=1,X=x_j]}{\Pr[D=0,X=x_j]\Pr[S=1,D=1,X=x_j]})\}$. Simplifying, and by assumption C, $\widehat{p} = 1 - \sum_{j=1}^J \frac{\Pr[S=1,D=0,X=x_j]\Pr[D=1]}{\Pr[S=1,D=1]\Pr[D=0]} = 1 - \frac{\Pr[S=1,D=0]\Pr[D=1]}{\Pr[S=1,D=1]\Pr[D=0]} = 1 - \frac{\Pr[S=1|D=0]}{\Pr[S=1|D=1]} = p$.

Now, it is true that $n^*(y) = \sum_{j=1}^J \Pr[X = x_j|S = 1, D = 1] \frac{1-p_j}{1-p} g(y|x_j)$. Using definitions of p and p_j, this is equal to $\sum_{j=1}^J \{\Pr[X = x_j|S = 1, D = 1] \frac{\frac{\Pr\left[S=1,D=0,X=x_j\right]\Pr\left[D=1,X=x_j\right]}{\Pr\left[D=0,X=x_j\right]\Pr\left[S=1,D=1,X=x_j\right]}}{\frac{\Pr[S=1,D=0]\Pr[D=1]}{\Pr[D=0]\Pr[S=1,D=1]}} g(y|x_j)\}$. Applying Assumption C, this becomes $\sum_{j=1}^J \Pr[X = x_j|S = 1, D = 1] \frac{\Pr[S=1,D=0,X=x_j]\Pr[S=1,D=1]}{\Pr[S=1,D=1,X=x_j]\Pr[S=1,D=0]} g(y|x_j)$. Simplifying further yields $\sum_{j=1}^J \frac{\Pr[X=x_j,S=1,D=0]}{\Pr[S=1,D=0]} g(y|x_j) = \sum_{j=1}^J \Pr[X = x_j|S = 1, D = 0] g(y|x_j)$.

Now consider the case $\widehat{p} = p = 0$. This means $\Pr[S = 1|D = 1] = \Pr[S = 1|D = 0]$. $0 = \Pr[S = 1|D = 1] - \Pr[S = 1|D = 0] = \sum_j^J \{\Pr[X = x_j|D = 0] (\Pr[S = 1|D = 1, X = x_j] - \Pr[S = 1|\ D = 0, X = x_j])\}$ because of Assumption C. Since $\Pr[X = x_j|D = 0] > 0$ for all $j = 1, \ldots, J$, and Assumption B implies that $\Pr[S = 1|D = 1, X = x_j] - \Pr[S = 1|D = 0, X = x_j] \geq 0$ for $j = 1, \ldots, J$, then it must be true that $\Pr[S=1|D=1,X=x_j]-\Pr[S=1|D=0,X=x_j]=0$ for $j=1,\ldots,J$, which means that $p_j=0$ for $j=1,\ldots,J$. So no trimming is done at either the aggregate level or by values of X. It can then be shown, using Assumption C and Bayes' Rule, that this implies $\Pr[X=x_j|S=1,D=1]=\Pr[X=x_j|S=1,D=0]$ for $j=1,\ldots,J$. Then $\overline{E}\equiv\int_{-\infty}^{\infty}yf(y)\,dy=\sum_{j=1}^{J}\{\Pr[X=x_j|S=1,D=1]\int_{-\infty}^{\infty}yf(y|x_j)\,dy\}=\sum_{j=1}^{J}\Pr[X=x_j|S=1,D=0]\int_{-\infty}^{\infty}yf(y|x_j)\,dy=\overline{E}^{*}$.

References

[1] Andrews, D., and Schafgans, M. (1998), "Semiparametric Estimation of the Intercept of a Sample Selection Model," *Review of Economic Studies*, 65, 497-517.

[2] Ahn, H. and Powell, J. (1993), "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3-29.

[3] Angrist, J., Imbens, G., and Rubin, D. (1996) "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444-445.

[4] Angrist, J., (1997) "Conditional Independence in Sample Selection Models," *Economics Letters*, 54, 103-112.

[5] Balke, A., and Pearl, J. (1997), "Bounds on Treatment Effects from Studies With Imperfect Compliance," *Journal of the American Statistical Association*, 92, 1171-1177.

[6] Bloom, H. (1984), "Accounting for No-Shows in Experimental Evaluation Designs," *Evaluation Review*, 8, 225-246.

[7] Das, M., Newey, W. K., and Vella, F. (2000), "Nonparametric Estimation of Sample Selection Models", mimeo.

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

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

[10] Heckman, J. J. and R. 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* (Springer, New York).

[11] Heckman, J. J. (1990), "Varieties of Selection Bias," *American Economic Review Papers and Proceedings*, 80, 313-318.

[12] Heckman, J. J., and Vytlacil, E., (1999), "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96, 4730-4734.

[13] Heckman, J. J., and Vytlacil, E., (2000a), "Local Instrumental Variables," *National Bureau of Economic Research Technical Working Paper #252*.

[14] Heckman, J. J., and Vytlacil E., (2000b), "Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effect," *National Bureau of Economic Research Technical Working Paper #259*.

[15] Horowitz, J. L., and Manski, C. F. (1998), "Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations," *Journal of Econometrics*, 84, 37-58.

[16] Horowitz, J. L., and Manski, C. F. (2000a) "Nonparametric Analysis of Randomized Experiments With Missing Covariate and Outcome Data" *Journal of the American Statistical Association*, 95, 77-84.

[17] Horowitz, J. L., and Manski, C. F. (2000b) Rejoinder: "Nonparametric Analysis of Randomized Experiments With Missing Covariate and Outcome Data" *Journal of the American Statistical Association*, 95, 87.

[18] Imbens, G., and Angrist, J. (1994), "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, 62 (4): 467-476.

[19] Manski, C. F. (1989), "Anatomy of the Selection Problem," *Journal of Human Resources*, 24, 343-360.

[20] Manski, C. F. (1990), "Nonparametric Bounds on Treatment Effects," *American Economic Review Papers and Proceedings*, 80, 319-323.

[21] Manski, C. F. (1995), *Identification Problems in the Social Sciences*, Cambridge, MA: Harvard Uni-

versity Press.

[22] Robins, J. (1989), "The Analysis of Randomized and Non-Randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies," in *Health Service Research Methodology: A Focus on AIDS*, eds. L. Sechrest, H. Freeman, and A. Mulley, Washington, DC: U.S. Public Health Service.

[23] Rubin, D. (1976), "Inference and Missing Data" *Biometrika*, 63, 581-592.

[24] Vytlacil, E. (2000), "Independence, Monotonicity, and Latent Index Models: An Equivalence Result" *mimeo*.