I Z A

IZA DP No. 3632

Minimizing Bias in Selection on Observables Estimators When Unconfoundness Fails

Daniel L. Millimet
Rusty Tchernis

August 2008

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Minimizing Bias in Selection on Observables Estimators When Unconfoundness Fails

Daniel L. Millimet
*Southern Methodist University
and IZA*

Rusty Tchernis
*Indiana University
and NBER*

ABSTRACT

Minimizing Bias in Selection on Observables Estimators When Unconfoundness Fails[*]

We characterize the bias of propensity score based estimators of common average treatment effect parameters in the case of selection on unobservables. We then propose a new minimum biased estimator of the average treatment effect. We assess the finite sample performance of our estimator using simulated data, as well as a timely application examining the causal effect of the School Breakfast Program on childhood obesity. We find our new estimator to be quite advantageous in many situations, even when selection is only on observables.

Corresponding author:

Daniel L. Millimet
Department of Economics
Southern Methodist University
Box 0496
Dallas, TX 75275-0496
USA
E-mail: millimet@mail.smu.edu

---

1   Introduction

Program evaluation methods in econometrics have become increasingly sophisticated. In this study, we assess and compare the bias of estimators of three common parameters – the average treatment effect (ATE), the average treatment effect on the treated (ATT), and the average treated on the untreated (ATU) – that *require* the unconfoundedness assumption when, in fact, unconfoundedness *fails*. Thus, our study represents an extension to the analysis first put forth in Black and Smith (2004) for the ATT. In that paper, the authors show that under certain conditions, estimators of the ATT that require the unconfoundedness assumption to be unbiased, but are biased when this assumption fails to hold, obtain a minimum bias when the sample is restricted to observations with a probability of treatment, conditional on covariates, close to one-half.

In this paper, we have four specific objectives. First, we assess the bias of so-called selection on observables estimators of the ATE and ATU under the same assumptions as in Black and Smith (2004). Second, we provide a set of assumptions – frequently invoked in applied settings – that enable us to rank the magnitude of the bias of estimators across the three parameters. Third, we propose a new estimation technique for researchers seeking to minimize the bias of selection on observables estimators of the ATE when unconfoundedness fails to hold. We then compare the finite sample performance of our proposed method by Monte Carlo methods. Finally, we compare the performance of our estimator to those currently utilized in the literature to assess the effects of participation in the national School Breakfast Program (SBP) on childhood obesity.

Our results should be of interest to the growing number of applied researchers relying on estimators from the program evaluation literature, especially when the researcher is concerned that the unconfoundedness assumption is problematic but a valid exclusion restriction is unavailable. Specifically, we offer two recommendations to applied researchers. First and foremost, researchers ought to be skeptical of estimates obtained using parametric methods unless the correct functional form is known. Second, our new estimation technique represents an improvement over the performance of the semiparametric estimator of the ATE proposed in Hirano and Imbens (2001). In fact, our bias minimizing approach yields a finite sample improvement over Hirano and Imbens (2001) even when the underlying assumptions of their estimator are valid in the population. However, our estimation technique is particularly useful when there is selection on unobservables. While the fact that our estimator is preferable even when unconfoundedness holds may appear striking at first glance, it is consistent with other results in the literature. In particular, Hirano et al. (2003) find that using an estimated propensity score is preferable even when the true propensity score is known, and Millimet and Tchernis (2007) find that over-fitting the propensity score equation is

preferable to using the true functional form.

The remainder of the paper is organized as follow. Section 2 begins by providing a quick overview of the potential outcomes and corresponding treatment effects framework. Next, it contains our analysis of the bias of each of the three parameters considered herein under certain assumptions. Finally, we propose a new estimation method in Section 2. Our estimator attempts to minimize the bias of estimators of the ATE that rely on unconfoundedness when, in fact, this assumption is false. Section 3 presents a Monte Carlo study comparing the performance of our proposed estimator to other estimators commonplace in the literature. Section 4 contains the application to school nutrition programs. Section 5 concludes.

2   The Evaluation Problem

2.1   Setup

Consider a random sample of N individuals from a large population indexed by $i = 1, ..., N$. Utilizing the potential outcomes framework (see, e.g., Neyman 1923; Fisher 1935; Roy 1951; Rubin 1974), let $Y_i(T)$ denote the potential outcome of individual i under treatment T, $T \in \mathcal{T}$. Here, we consider only the case of binary treatments: $\mathcal{T} = \{0, 1\}$. The causal effect of the treatment $(T = 1)$ relative to the control $(T = 0)$ is defined as the difference between the corresponding potential outcomes. Formally,

$$\tau_i = Y_i(1) - Y_i(0). \tag{1}$$

In the evaluation literature, several population parameters are of potential interest. The most commonly used include the ATE, the ATT, and the ATU. These are defined as

$$\tau_{ATE} \;\; = \;\; \mathrm{E}[\tau_i] = \mathrm{E}[Y_i(1) - Y_i(0)] \tag{2}$$

$$\tau_{ATT} \;\; = \;\; \mathrm{E}[\tau_i|T = 1] = \mathrm{E}[Y_i(1) - Y_i(0)|T = 1] \tag{3}$$

$$\tau_{ATU} \;\; = \;\; \mathrm{E}[\tau_i|T = 0] = \mathrm{E}[Y_i(1) - Y_i(0)|T = 0]. \tag{4}$$

In general, the parameters in $(2) - (4)$ may vary with a vector covariates, X. As a result, each of the parameters may be defined conditional on a particular value of X as follows:

$$\tau_{ATE}[X] \;\; = \;\; \mathrm{E}[\tau_i|X] = \mathrm{E}[Y_i(1) - Y_i(0)|X] \tag{5}$$

$$\tau_{ATT}[X] \;\; = \;\; \mathrm{E}[\tau_i|X, T = 1] = \mathrm{E}[Y_i(1) - Y_i(0)|X, T = 1] \tag{6}$$

$$\tau_{ATU}[X] \;\; = \;\; \mathrm{E}[\tau_i|X, T = 0] = \mathrm{E}[Y_i(1) - Y_i(0)|X, T = 0]. \tag{7}$$

The parameters in (2) – (4) are obtained by taking the expectation of the corresponding parameter in (5) – (7) over the distribution of X in the relevant population (the unconditional distribution of X for the ATE, and distribution of X condition on $T = 1$ and $T = 0$ for the ATT and ATU, respectively).

For each individual, we observe the triple $\{Y_i, T_i, X_i\}$, where Y_i is the observed outcome, T_i is a binary indicator of the treatment received, and X_i is a vector of covariates. The only requirement of the covariates included in X_i is that they are pre-determined (that is, they are unaffected by T_i) and do not perfectly predict treatment assignment. The relationship between the potential and observed outcomes is given by

$$Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0) \tag{8}$$

which makes clear that only one potential outcome is observed for any individual. As such, estimating τ is not trivial as there is an inherent missing data problem; some assumptions are required to proceed.

One such assumption is *unconfoundedness* or selection on observables (Rubin 1974; Heckman and Robb 1985). Under this assumption, treatment assignment is said to be independent of potential outcomes conditional on the set of covariates, X. As a result, selection into treatment is random conditional on X and the average effect of the treatment can be obtained by comparing outcomes of individuals in different treatment states with identical values of the covariates. To solve the dimensionality problem that is likely to arise if X is a lengthy vector, Rosenbaum and Rubin (1983) propose using the propensity score, $P(X_i) = \Pr(T_i = 1|X_i)$, instead of X as the conditioning variable.

2.2   Bias When Unconfoundedness Fails

Given knowledge of the propensity score, or an estimate thereof, and sufficient overlap between the distributions of the propensity score across the $T = 1$ and $T = 0$ groups (typically referred to as the *common support* condition; see Dehejia and Wahba (1999) or Smith and Todd (2005)), the parameters discussed above can be estimated in a number of ways under the unconfoundedness assumption (D'Agostino (1998) and Imbens (2004) offer summaries). Regardless of which such technique is employed, each will be biased if the unconfoundedness assumption fails to hold.

Black and Smith (2004) consider the bias when estimating the ATT under unconfoundedness and the

assumption is incorrect. The bias of the ATT at some value of the propensity score, $P(X)$, is given by

$$
\begin{aligned}
B_{ATT}[P(X)] &= \widehat{\tau}_{ATT}[P(X)] - \tau_{ATT}[P(X)] \\
&= \{\mathrm{E}[Y(1)|T=1,P(X)] - \mathrm{E}[Y(0)|T=0,P(X)]\} \\
&\quad - \{\mathrm{E}[Y(1)|T=1,P(X)] - \mathrm{E}[Y(0)|T=1,P(X)]\} \\
&= \mathrm{E}[Y(0)|T=1,P(X)] - \mathrm{E}[Y(0)|T=0,P(X)] \quad (9)
\end{aligned}
$$

where $\widehat{\tau}_{ATT}$ refers some propensity score based estimator of the ATT (e.g., propensity score matching or inverse propensity score weighting).

To better understand the behavior of the bias, Black and Smith (2004) make the following two assumptions:

(A1) Potential outcomes and latent treatment assignment are additively separable in observables and unobservables

$$
\begin{aligned}
Y(0) &= g_0(X) + \varepsilon_0 \\
Y(1) &= g_1(X) + \varepsilon_1 \\
T^* &= h(X) - u \\
T &= \begin{cases} 1 & \text{if } T^* > 0 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

(A2) $\varepsilon_0, \varepsilon_1, u \sim N_3(0, \Sigma)$, where

$$
\Sigma = \begin{bmatrix} \sigma_0^2 & \rho_{01} & \rho_{0u} \\ & \sigma_1^2 & \rho_{1u} \\ & & 1 \end{bmatrix}.
$$

Given A1 and A2, (9) simplifies to

$$
\begin{aligned}
B_{ATT}[P(X)] &= \mathrm{E}[\varepsilon_0|T=1,P(X)] - \mathrm{E}[\varepsilon_0|T=0,P(X)] \\
&= \rho_{0u}\sigma_0 \frac{\phi(h(X))}{\Phi(h(X))[1-\Phi(h(X))]} \quad (10)
\end{aligned}
$$

where $\phi(\bullet)$ and $\Phi(\bullet)$ are the standard normal density and cumulative density function, respectively. As noted in Black and Smith (2004), $B_{ATT}[P(X)]$ is minimized when $h(X) = 0$, which implies that $P(X) =$

0.5. Thus, the authors recommend that researchers estimate τ_{ATT} using the 'thick support' region of the propensity score (e.g., $P(X) \in (0.33, 0.67)$).

Prior to continuing, it is important to note that if the ATT varies with X (and, hence, $P(X)$), then using only observations on the thick support estimates a different parameter than the population ATT given in (3). Indeed, the procedure suggested in Black and Smith (2004) accomplishes the following. It searches over the parameters defined in (6) to find the value of $P(X)$ for which the $\tau_{ATT}[P(X)]$ can be estimated with the least bias. This subtle, or perhaps not so subtle, point is very intriguing. Stated differently, when unconfoundedness fails, τ_{ATT}, the population ATT, cannot be estimated in an unbiased manner using estimators that rely on this assumption. Rather than invoking different assumptions to identify the population ATT (e.g., those utilized by selection on unobservables estimators), the Black and Smith (2004) approach identifies the *parameter that can be estimated with the smallest bias under unconfoundedness.* Whether or not the parameter being estimated with the least bias, $\widetilde{\tau}_{ATT} = \mathrm{E}[\mathrm{E}[\tau_i | P(X), T = 1]]$, where the outer expectation is over $X | 0.33 < P(X) < 0.67$ and $T = 1$, is an interesting economic parameter is a different question. The key point, however, is that when restricting the estimation sample to observations with propensity scores contained in a subset of the unit interval, the parameter being estimated will differ from the population ATT unless the treatment effect does not vary with X (i.e., $\mathrm{E}[\tau_i | P(X), T = 1] = \mathrm{E}[\tau_i | T = 1]$).

With this point in mind, we now consider the bias for the ATE and the ATU since the ATT is not the only parameter of interest in applied settings. For the ATU, it is trivial to show that

$$
\begin{aligned}
B_{ATU}[P(X)] &= \mathrm{E}[\varepsilon_1 | T = 1, P(X)] - \mathrm{E}[\varepsilon_1 | T = 0, P(X)] \\
&= \rho_{1u}\sigma_1 \frac{\phi(h(X))}{\Phi(h(X))[1 - \Phi(h(X))]}
\end{aligned}
\tag{11}
$$

which is also minimized when $h(X) = 0$, or $P(X) = 0.5$. However, it is useful to note that

$$
\begin{aligned}
B_{ATU}[P(X)] &= \mathrm{E}[\delta + \varepsilon_0 | T = 1, P(X)] - \mathrm{E}[\delta + \varepsilon_0 | T = 0, P(X)] \\
&= B^{ATT}[P(X)] + \mathrm{E}[\delta | T = 1, P(X)] - \mathrm{E}[\delta | T = 0, P(X)]
\end{aligned}
$$

where $\delta = \varepsilon_1 - \varepsilon_0$ is the unobserved, individual-specific gain from treatment. Thus, the magnitude of the bias of the ATU may either be larger or smaller than the corresponding bias of the ATT. If we add the following assumption:

(A3) Non-negative selection into the treatment on individual-specific, unobserved gains

$$\mathrm{E}[\delta|T = 1, P(X)] \geqslant \mathrm{E}[\delta|T = 0, P(X)]$$

then $B_{ATU}[P(X)] \geqslant B_{ATT}[P(X)]$ for all $P(X)$.[1]

Now consider the ATE. Utilizing the fact that $\tau_{ATE}[P(X)] = P(X)\tau_{ATT}[P(X)]+[1-P(X)]\tau_{ATU}[P(X)]$, and rewriting $Y(1) = g_1(X) + (\delta + \varepsilon_0)$, the bias for the ATE is given by

$$
\begin{aligned}
B_{ATE}[P(X)] &= \rho_{0u}\sigma_0 \left\{ \frac{\phi(h(X))}{\Phi(h(X))[1 - \Phi(h(X))]} \right\} + [1 - P(X)] \left\{ \rho_{\delta u}\sigma_\delta \frac{\phi(h(X))}{\Phi(h(X))[1 - \Phi(h(X))]} \right\} \\
&= \{\rho_{0u}\sigma_0 + [1 - P(X)]\rho_{\delta u}\sigma_\delta\} \left\{ \frac{\phi(h(X))}{\Phi(h(X))[1 - \Phi(h(X))]} \right\}. \tag{12}
\end{aligned}
$$

Equation (12) leads to three salient points. First, the bias of the ATE is minimized when $h(X) = 0$ only in the case where $\rho_{\delta u} = 0$ (i.e., no selection on unobserved, individual-specific gains to treatment). Second, under A1 – A3, $B_{ATU}[P(X)] \geqslant B_{ATE}[P(X)] \geqslant B_{ATT}[P(X)]$. Third, the value of $P(X)$ that minimizes the bias of the ATE is not fixed; rather, it depends on the signs and magnitudes of $\rho_{0u}\sigma_0$ and $\rho_{\delta u}\sigma_\delta$. Simulations (see Figures 1 and 2) reveal that $B_{ATE}[P(X)]$ displays the following properties:

(i) if $\rho_{0u}\sigma_0 = 0$, then $\lim_{P(X)\longrightarrow 1} B_{ATE}[P(X)] = 0$

(ii) if $\rho_{0u}\sigma_0 = -\rho_{\delta u}\sigma_\delta$, then $\lim_{P(X)\longrightarrow 0} B_{ATE}[P(X)] = 0$

(iii) if $\rho_{0u}\sigma_0 > 0$, then $P^* = \arg\min B_{ATE}[P(X)]$ is

    (a) strictly greater than 0.5

    (b) decreasing in $\rho_{0u}\sigma_0$ (holding $\rho_{\delta u}\sigma_\delta$ fixed)

    (c) increasing in $\rho_{\delta u}\sigma_\delta$ (holding $\rho_{0u}\sigma_0$ fixed)

(iv) if $\rho_{0u}\sigma_0 < 0$, then $P^* = \arg\min B_{ATE}[P(X)]$ is

    (a) above or below 0.5, but monotonically increasing, for $\rho_{0u}\sigma_0 \in (-\rho_{\delta u}\sigma_\delta, 0)$

    (b) strictly less than 0.5, but decreasing in $\rho_{0u}\sigma_0$ (holding $\rho_{\delta u}\sigma_\delta$ fixed) for $\rho_{0u}\sigma_0 < -\rho_{\delta u}\sigma_\delta$

    (c) decreasing in $\rho_{\delta u}\sigma_\delta$ (holding $\rho_{0u}\sigma_0$ fixed) for $\rho_{0u}\sigma_0 < -\rho_{\delta u}\sigma_\delta$.

where P^* is the value of the propensity score that minimizes the bias in (12). We refer to P^* as the *bias minimizing propensity score* (BMPS). See Figure 3 for a complete characterization of P^*.

---

[1]See Heckman and Vytlacil (2005) for further discussion on this assumption.

2.3 Estimation

To assess the bias of selection on observables estimators of the ATE when unconfoundedness fails, we utilize the normalized inverse probability weighted estimator of Hirano and Imbens (2001). Horvitz and Thompson (1952) show that the ATE may be expressed as

$$\tau_{ATE} = \mathrm{E}\left[\frac{Y \cdot T}{P(X)} - \frac{Y \cdot (1-T)}{1 - P(X)}\right], \tag{13}$$

with the sample analogue given by

$$\hat{\tau}_{HT} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{Y_i T_i}{\hat{P}(X_i)} - \frac{Y_i(1-T_i)}{1 - \hat{P}(X_i)}\right]. \tag{14}$$

The estimator in (14) is the *unnormalized* estimator as the weights do not necessarily sum to unity. To circumvent this issue, Hirano and Imbens (2001) propose an alternative estimator, referred to as the *normalized* estimator, which is given by

$$\hat{\tau}_{HI} = \left[\sum_{i=1}^{N}\frac{Y_i T_i}{\hat{P}(X_i)}\Bigg/\sum_{i=1}^{N}\frac{T_i}{\hat{P}(X_i)}\right] - \left[\sum_{i=1}^{N}\frac{Y_i(1-T_i)}{1 - \hat{P}(X_i)}\Bigg/\sum_{i=1}^{N}\frac{(1-T_i)}{1 - \hat{P}(X_i)}\right]. \tag{15}$$

Millimet and Tchernis (2007) provide evidence of the superiority of the normalized estimator in practical settings.

Under unconfoundedness, the normalized estimator in (15) provides an unbiased estimate of τ_{ATE}. When this assumption fails, the bias is given (12). To minimize the bias, we propose to estimate (15) using only observations with a propensity score in a neighborhood around the BMPS, P^*. Formally, we propose the following *minimum biased* estimator of the ATE:

$$\hat{\tau}_{MB}[P^*] = \left[\sum_{i\in\Omega}\frac{Y_i T_i}{\hat{P}(X_i)}\Bigg/\sum_{i\in\Omega}\frac{T_i}{\hat{P}(X_i)}\right] - \left[\sum_{i\in\Omega}\frac{Y_i(1-T_i)}{1 - \hat{P}(X_i)}\Bigg/\sum_{i\in\Omega}\frac{(1-T_i)}{1 - \hat{P}(X_i)}\right] \tag{16}$$

where

$$\Omega = \{i|\hat{P}(X_i) \in C(P^*)\},$$

and $C(P)$ denotes a neighborhood around P. In the estimation below, we define $C(P^*)$ as

$$C(P^*) = \{\hat{P}(X_i)|\hat{P}(X_i) \in (\underline{P}, \overline{P})\},$$

where $\underline{P} = \max\{0.02, P^* - \alpha_\theta\}$, $\overline{P} = \min\{0.98, P^* + \alpha_\theta\}$, and $\alpha_\theta > 0$ is the smallest value such that at

least θ percent of both the treatment and control groups are contained in Ω. In the exercises below, we set $\theta = 0.05$, 0.10, and 0.25. For example, if $\theta = 0.05$, we find the smallest value, $\alpha_{0.05}$, such that 5% of the treatment group and 5% of the control group have a propensity score in the interval $(\underline{P}, \overline{P})$. Thus, smaller values of θ should reduce the bias at the expense of higher variance. Note, we trim observations with propensity scores above (below) 0.98 (0.02), regardless of the value of θ, to prevent any single observations from receiving too large of a weight.

As defined above, the set Ω is unknown since, in general, P^* is unknown. To estimate the set Ω, we propose to estimate P^* assuming A1, A2, and functional forms for $g_0(X)$, $g_1(X)$, and $h(X)$ using the Heckman bivariate normal (BVN) selection model. Specifically, assuming

$$
\begin{aligned}
g_0(X) &= X\beta_0 \\
g_1(X) &= X\beta_1 \\
h(X) &= X\gamma
\end{aligned}
$$

then

$$
y_i = X_i\beta_0 + X_iT_i(\beta_1 - \beta_0) + \beta_{\lambda 0}(1 - T_i)\left[\frac{-\phi(X_i\gamma)}{1 - \Phi(X_i\gamma)}\right] + \beta_{\lambda 1}T_i\left[\frac{\phi(X_i\gamma)}{\Phi(X_i\gamma)}\right] + \eta_i \tag{17}
$$

where $\phi(\cdot)/\Phi(\cdot)$ is the inverse Mills' ratio, η is a well-behaved error term, and

$$
\begin{aligned}
\beta_{\lambda 0} &= \rho_{0u}\sigma_0 \\
\beta_{\lambda 1} &= \rho_{0u}\sigma_0 + \rho_{\delta u}\sigma_\delta.
\end{aligned} \tag{18}
$$

Thus, OLS estimation of (17) after replacing γ with an estimate obtained from a first-stage probit model yields consistent estimates of $\rho_{0u}\sigma_0$ and $\rho_{\delta u}\sigma_\delta$. With these estimates, one can use (12) to obtain an estimate of P^*.[2]

Our proposed estimator immediately raises a question: If one is willing to maintain the assumptions underlying the BVN selection model in (17), why not just use the OLS estimates of (17) to estimate the ATE? Perhaps one should. If the assumptions underlying the BVN selection model are valid, then one should; the ATE is estimated by

$$
\hat{\tau}_{BVN} = \overline{X}\left(\widehat{\beta}_1 - \widehat{\beta}_0\right). \tag{19}
$$

However, if these assumptions fail, then perhaps the estimator in (15) or (16) will perform better in practice.

---

[2] To estimate P^*, we conduct a grid search over 1,000 equally-spaced values of $h(\cdot)$ from -5 to 5. If P^* is above 0.98, we truncate it to 0.98; if P^* is below 0.02, we truncate it to 0.02.

To better understand the finite sample performance of these estimators, we now turn to our Monte Carlo study.

3   Monte Carlo Study

3.1   Setup

To assess the performance of the various estimators, we use five experimental designs. However, for each experimental design, we use two data-generating processes (DGPs): (i) a constant treatment effect setup (i.e., $\tau_i = \tau$ for all i), and (ii) a heterogeneous treatment effect setup, where the heterogeneity is due to unobserved, individual-specific gains to treatment (i.e., τ_i varies across i, but this variation is unrelated to variation in observables, X). Thus, we are focusing on cases where all of the estimators being compared are estimating the same underlying parameter, $\tau_{ATE} = \mathrm{E}[\tau_i]$.

Each experiment entails simulating 500 data sets containing

$$x_1, x_2 \overset{iid}{\sim} N(0, 4)$$

and

$$
\begin{aligned}
h(X) &= x_1 + x_2 - 0.5(x_1^2 - x_2^2) + x_1 x_2 \\
T^* &= h(X) - u \\
T &= \begin{cases} 1 & \text{if } T^* > 0 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

for 1000 observations. Furthermore, in the constant treatment effect setup:

$$
\begin{aligned}
Y(0) &= g_0(X) + \varepsilon_0 = h(X) + \varepsilon_0 \\
Y(1) &= g_1(X) + \varepsilon_0 = 1 + h(X) + \varepsilon_0
\end{aligned}
$$

which implies that $\tau_i = 1$ for all i and $\tau_{ATE} = \mathrm{E}[\tau_i] = 1$. In the heterogeneous treatment effect setup:

$$
\begin{aligned}
Y(0) &= g_0(X) + \varepsilon_0 = h(X) + \varepsilon_0 \\
Y(1) &= g_1(X) + \varepsilon_1 = 1 + h(X) + \varepsilon_1
\end{aligned}
$$

which implies that $\tau_i = 1 + \varepsilon_{1i} - \varepsilon_{0i} = 1 + \delta_i$ and $\tau_{ATE} = \mathrm{E}[\tau_i] = 1$ (assuming $\mathrm{E}[\delta_i] = 0$, as discussed

below).

Our five experimental designs are:

1. *All assumptions of the BVN selection model are correct.* In particular, in the constant treatment effect setup,

$$\varepsilon_0, u \sim N_2(0, \Sigma)$$

where

$$\Sigma = \begin{bmatrix} 1 & \rho_{0u} \\ & 1 \end{bmatrix}.$$

In the heterogeneous treatment effect setup,

$$\varepsilon_0, \varepsilon_1, u \sim N_3(0, \Sigma)$$

where

$$\Sigma = \begin{bmatrix} 1 & 0.5 & \rho_{0u} \\ & 1 & \rho_{1u} \\ & & 1 \end{bmatrix}.$$

In addition, in both cases, the correct functional forms of the first-stage propensity score equation and the second-stage outcome equation in (17) are assumed to be known.

2. *All assumptions of the BVN selection model are correct except the second-stage outcome equation in (17) is mis-specified.* Specifically, the error distributions are identical to experiment #1, and the first-stage propensity score equation is assumed to be known, but x_1^2, x_2^2, and $x_1 x_2$ are omitted from the set of covariates included in (17).

3. *All assumptions of the BVN selection model are correct except the first-stage propensity score equation and the second-stage outcome equation in (17) are mis-specified.* Specifically, the error distributions are identical to experiment #1, but x_1^2, x_2^2, and $x_1 x_2$ are omitted from the set of covariates included in both the first-stage propensity score equation and (17).

4. *All assumptions of the BVN selection model are correct except the errors are uniformly distributed.* In particular, in the constant treatment effect setup,

$$\varepsilon_0, u \sim U_2(0, \Sigma)$$

10

where

$$\Sigma = \begin{bmatrix} 1 & \rho_{0u} \\ & 1 \end{bmatrix}.$$

In the heterogeneous treatment effect setup,

$$\varepsilon_0, \varepsilon_1, u \sim U_3(0, \Sigma)$$

where

$$\Sigma = \begin{bmatrix} 1 & 0.5 & \rho_{0u} \\ & 1 & \rho_{1u} \\ & & 1 \end{bmatrix}.$$

In both cases, the correct functional forms of the first-stage propensity score equation and the second-stage outcome equation in (17) are assumed to be known.

5. *All assumptions of the BVN selection model are correct except the errors are uniformly distributed and the second-stage outcome equation in (17) is mis-specified.* Specifically, the error distributions are identical to experiment #4, and the first-stage propensity score equation is assumed to be known, but x_1^2, x_2^2, and $x_1 x_2$ are omitted from the set of covariates included in (17).

Each experiment is conducted for several combinations of true values for $\rho_{0u}\sigma_0$ and $\rho_{\delta u}\sigma_\delta$, including the case where unconfoundedness holds. To assess the performance of the estimators, we report the mean estimates of $\rho_{0u}\sigma_0$ and $\rho_{\delta u}\sigma_\delta$, and the mean values of α_θ, $\theta = 0.05$, 0.10, and 0.25. In addition, we report the mean bias, the mean absolute error (MAE), and the mean squared error (MSE) for each estimator. Note, for the constant treatment effect setup, we estimate ρ_{0u} and (19) using

$$y_i = X_i\beta_0 + X_i T_i(\beta_1 - \beta_0) + \beta_\lambda \left\{ (1 - T_i) \left[\frac{-\phi(X_i\gamma)}{1 - \Phi(X_i\gamma)} \right] + T_i \left[\frac{\phi(X_i\gamma)}{\Phi(X_i\gamma)} \right] \right\} + \eta_i \qquad (20)$$

where $\beta_\lambda = \rho_{0u}\sigma_0$.

3.2 Results

The results for the five experimental designs are presented in Tables 1-5. Table 1 is the benchmark case; the DGPs satisfy the assumptions of the BVN model, and the correct functional forms are utilized in the first-stage probit model for the propensity score and second-stage equation for the outcome. As mentioned above, even absent an exclusion restriction in the first-stage, we expect the BVN estimator to outperform the other estimators given the efficiency gain from imposing restrictions that hold in the DGP.

The results in Table 1 confirm our expectations. In all cases except one, $\widehat{\tau}_{BVN}$ has the smallest bias. In addition, $\widehat{\tau}_{BVN}$ has the smallest MAE and MSE in all cases, including the two cases where unconfoundedness holds ($\rho_{0u}\sigma_0 = 0$ in the constant treatment effect setup and $\rho_{0u}\sigma_0 = \rho_{\delta u}\sigma_\delta = 0$ in the heterogeneous treatment effect setup). However, if one restricts attention to only the estimators that require unconfoundedness for unbiasedness, our estimator, $\widehat{\tau}_{MB}$, outperforms the Hirano and Imbens (2001) estimator, $\widehat{\tau}_{HI}$, in all cases when there is at least some selection on unobservables. Moreover, even when selection on unobservables holds, our minimum biased estimator still outperforms $\widehat{\tau}_{HI}$ for at least some value of θ.

Among the minimum biased estimators, utilizing a smaller value of θ is preferable in terms of bias, MAE, and MSE in the majority of cases when selection on unobservables is 'sufficiently' large. However, a larger value of θ is preferable in terms of MSE under unconfoundedness or when selection on unobservables is 'sufficiently' weak. For example, $\widehat{\tau}_{MB,0.05}$ or $\widehat{\tau}_{MB,0.10}$ yields the smallest bias, MAE, and MSE when $\rho_{0u}\sigma_0 > 0$ and $\rho_{\delta u}\sigma_\delta > 0$ in the heterogeneous treatment effect setup; $\widehat{\tau}_{MB,0.05}$ or $\widehat{\tau}_{MB,0.10}$ yields the smallest bias and MAE in the constant treatment effect setup when $\rho_{0u}\sigma_0 > 0.5$ (but a higher MSE when $\rho_{0u}\sigma_0 = 0.25$). Conversely, when unconfoundedness holds in either the heterogeneous or constant treatment effect setup, $\widehat{\tau}_{MB,0.25}$ yields the smallest MAE and MSE. Thus, when using weighting estimators of the ATE, our minimum biased estimator is preferable on both bias and efficiency grounds relative to Hirano and Imbens (2001) estimator, and the value of θ should be decreasing in the amount of selection on unobservables suspected by the researcher.

Prior to continuing, it is worth emphasizing the point that the 'optimal' θ does not go to unity when unconfoundedness holds. Thus, even when selection on observables holds in the population, finite sample performance can be improved by removing any within sample correlation among the errors. This finding is consistent with the results in Hirano et al. (2003), who find that using an estimated propensity score is preferable even when the true propensity score is known, and Millimet and Tchernis (2007), who show that over-fitting the propensity score equation is preferable to using the true functional form.

Table 2 provides the results from our second experimental design. Recall, the only deviation from the design employed in Table 1 is that now the second-stage outcome equation in (17) and (20) is mis-specified; we omit x_1^2, x_2^2, and $x_1 x_2$ from the estimation. Thus, this corresponds to a situation where a researcher incorrectly assumes a linear functional form in the second-stage.

Note two important details before turning to the results. First, the first-stage probit model for the propensity score is specified correctly. As a result, the model now contains three exclusion restrictions in the probit equation, although they are not 'valid' exclusion restrictions since x_1^2, x_2^2, and $x_1 x_2$ also have a direct effect on y. Second, because the propensity score equation is correctly specified, the performance of

$\widehat{\tau}_{HI}$ will be identical to that in Table 1; the difference in the experimental designs does not impact this estimator. However, our estimator, $\widehat{\tau}_{MB}$, is affected because the estimate of the BMPS, P^*, depends on the estimates from the second-stage equation in (17) and (20). Thus, whether our estimator continues to outperform $\widehat{\tau}_{HI}$ is not clear *a priori*.

In terms of the results, three findings stand out. First, $\widehat{\tau}_{BVN}$ does *significantly* worse than any of the other estimators in terms of bias, MAE, and MSE. Specifically, the MAE ranges from roughly ten to 100 times as large; the MSE is 100 to 1000 times as large. Second, our estimator, $\widehat{\tau}_{MB}$, continues to outperform $\widehat{\tau}_{HI}$ in all cases. Thus, our estimator continuous to constitute an improvement upon $\widehat{\tau}_{HI}$ even when unconfoundedness holds (for at least some value(s) of θ).

Finally, while not universal, a smaller value θ is generally preferable in terms of bias, MAE, and MSE as the degree of selection of unobservables increases; $\widehat{\tau}_{MB,0.25}$ yields the smallest MAE only when unconfoundedness holds or when $\rho_{0u}\sigma_0 = 0$ and $\rho_{\delta u}\sigma_\delta = 0.15$ and the smallest MSE only in two other cases where the extent of selection of unobservables is not 'sufficiently' large (the constant treatment effect setup with $\rho_{0u}\sigma_0 = 0.25$ and the heterogeneous treatment effect setup with $\rho_{0u}\sigma_0 = 0.15$ and $\rho_{\delta u}\sigma_\delta = 0$). Thus, as the extent of selection on unobservables gets stronger, one wants to restrict the estimation sample to observations closer to the estimated BMPS, P^*.

Table 3 presents the results for the same DGPs as in the two preceding tables except now both the first-stage probit model and the second-stage outcome equation are mis-specified. Specifically, both are estimated incorrectly assuming a linear functional form; x_1^2, x_2^2, and $x_1 x_2$ are omitted. Unfortunately, this might be the most relevant case for applied researchers. The results are easily summarizable. First, none of the estimators perform well, but $\widehat{\tau}_{BVN}$ does *significantly* worse than the other estimators. Second, $\widehat{\tau}_{MB,0.05}$ yields the lowest bias, MAE, and MSE in all cases. Relative to $\widehat{\tau}_{MB,0.05}$, the MAE (MSE) of $\widehat{\tau}_{HI}$ is roughly two to three (four to ten) times larger.

The results in Table 3 should have been expected given the prior results in Table 2. The common omitted covariates from both the first- and second-stage equations generate a strong correlation between the error terms in the potential outcome equation in the untreated state and treatment assignment equation even if $\rho_{0u}\sigma_0 = 0$ in the population. Thus, given the mis-specification, all cases in experimental design correspond to the case of 'sufficiently' strong selection on unobservables. In this case, as we saw in Table 2, $\theta = 0.05$ is preferable.

Table 4 displays the results from the next experimental design. Here, we revert back to the benchmark case in that we assume the correct functional form for both the first-stage propensity score model and second-stage outcome equation are known. However, the error terms are no longer normally distributed; rather, they are drawn from a correlated multivariate uniform distribution. The results are interesting.

First, $\widehat{\tau}_{BVN}$ clearly dominates the other estimators, even when unconfoundedness holds. Presumably this is because the (incorrect) normality assumption invoked in the first-stage probit model is a reasonable approximation.[3] Second, our estimator continues to outperform $\widehat{\tau}_{HI}$ when unconfoundedness holds (for $\theta = 0.10$ or 0.25), and when it does not (for all values of θ).

Third, as in the previous cases, within our estimator a smaller value θ is preferable in terms of MAE and MSE as the degree of selection of unobservables increases. For instance, $\widehat{\tau}_{MB,0.25}$ and $\widehat{\tau}_{MB,0.10}$ yield the smallest MAE and MSE (within our estimator) in the constant treatment effect setup when unconfoundedness holds, but $\widehat{\tau}_{MB,0.10}$ is preferable in the other two cases. In the heterogeneous treatment effect setup, $\widehat{\tau}_{MB,0.25}$ yields the smallest MAE and MSE (within our estimator) when $\rho_{0u}\sigma_0 = 0$ and $\rho_{\delta u}\sigma_\delta = 0$ or 0.25, or when $\rho_{0u}\sigma_0 = 0.15$ and $\rho_{\delta u}\sigma_\delta = 0$. $\widehat{\tau}_{MB,0.05}$ yields the smallest MAE and MSE (within our estimator) when $\rho_{0u}\sigma_0 = 0.30$ and $\rho_{\delta u}\sigma_\delta = 0.50$. Thus, there is essentially a negative, monotonic relationship between the 'optimal' θ and the extent of selection on unobservables. Specifically, $\theta = 0.25$ minimizes the MAE and MSE (relative to the other values of θ employed) when unconfoundedness holds; $\theta = 0.05$ minimizes the MAE and MSE when selection on unobservables is 'sufficiently' strong; $\theta = 0.10$ does best when selection on unobservables is present, but a bit 'weaker.'

Our final experimental design combines the error mis-specification from the previous DGP, with the functional form mis-specification of the second-stage outcome equation analyzed in the second experiment. The results are provided in Table 5. Again, several interesting findings emerge. First, as in Table 2, $\widehat{\tau}_{BVN}$ performs *significantly* worse than the other estimators. Thus, while error mis-specification (of the type analyzed here) is not particularly problematic for the BVN estimator, mis-specification of the outcome equation is very troublesome. Second, our estimator continues to outperform $\widehat{\tau}_{HI}$ when unconfoundedness holds (for $\theta = 0.10$ and 0.25), and when it does not (for all values of θ). In the heterogeneous treatment effect setup under unconfoundedness, the MAE and MSE of $\widehat{\tau}_{HI}$ is 1.3 and 1.5 times higher, respectively, than of $\widehat{\tau}_{MB,0.25}$. Finally, there continues to exist a negative, monotonic relationship between the 'optimal' θ and the extent of selection on unobservables as in Table 4.

3.3 Discussion

The simulation results presented, while clearly not exhausting all situations applied researchers confront, suggest a number of guidelines that researchers ought to find useful. First and foremost, researchers ought to be very wary when using the BVN estimator. Aside from the reliance on distributional assumptions, our study indicates that utilizing the proper functional form in the outcome equation is crucial. It is *so crucial* that we obtain drastically better estimates utilizing semiparametric estimators that *require*

---

[3]In test runs using our DGP, we do, however, reject normality of u using the test developed in Royston (1991).

unconfoundedness even when unconfoundedness *does not hold*.

Second, regardless of whether the outcome equation used in the BVN model is correctly specified, restricting the sample to only those observations whose propensity score lies in a certain region of the unit interval, where the restricted region is determined utilizing estimates obtained from the estimation of the BVN model, improves the performance of the semiparametric estimator proposed in Hirano and Imbens (2001). Thus, our bias minimizing approach yields a finite sample improvement over Hirano and Imbens (2001) even when the underlying assumptions of that estimator are valid in the population.

Finally, how 'tight' to restrict the sample used in the estimation depends on the underlying correlation structure of the unobservables in the model. Among the cases considered here, the 'optimal' restriction becomes tighter as the degree of selection on unobservables increases. When unconfoundedness holds, utilizing $\theta = 0.25$ achieves the best performance of the values considered here. However, when selection on unobservables is present, but is 'modest' in some sense, utilizing $\theta = 0.10$ tends to do best, while $\theta = 0.05$ is preferable under relatively 'strong' selection on unobservables.

In light of these recommendations, one might wonder why restricting the sample to an interval determined by estimates obtained from the BVN model, even when the BVN model is mis-specified and the resulting estimates seem quite biased, yields an improvement over the Hirano and Imbens (2001) estimator. Do our results indicate that restricting the estimation sample to observations with a propensity score lying in *some* subset of the unit interval would improve on the finite sample performance of the Hirano and Imbens (2001) estimator?

Tables 6-10 get at this question a bit. Specifically, we compare the results of our minimum biased estimator obtained in Tables 1-5 with those obtained fixing the BMPS, P^*, at its true value (i.e., the optimal value using the values of $\rho_{0u}\sigma_0$ and $\rho_{\delta u}\sigma_\delta$ from the DGP).[4] The results indicate three findings. First, when the parametric assumptions are all valid (Table 6) or only the error distribution is mis-specified (Table 9), there is little substantive difference in performance between using our estimated P^* and the true P^*. This is not surprising given our previous finding that $\widehat{\tau}_{BVN}$ performs well in this two experiments. Second, when both the outcome equation is mis-specified with or without the error distribution also being mis-specified (Tables 7 and 10), the is a definite gain to knowing the true P^*. Note, this is a not an indictment of our estimation procedure, but it does show that the performance of our estimator, albeit better than the Hirano and Imbens (2001) estimator, can do even better with an improved estimate of P^*. Finally, when both the propensity score equation and the outcome equation are mis-specified (Table 7), using the true P^* does *significantly* worse relative to using our estimated P^*. As a result, when we

---

[4]Note, when unconfoundness holds, P^* does not take on a unique value, but all values of P^* minimize the bias (and the bias is zero). Thus, we focus only on the cases where unconfoundedness does not hold.

are using poor estimates of the propensity score – because the equation is mis-specified – we are better off choosing the estimation sample using the estimated P^* based on the same mis-specified model, rather than choosing the estimation sample using the true P^* combined with the poorly estimated propensity scores.

In sum, then, our minimum biased estimator proves to be beneficial when unconfoundedness holds, but especially when it does not. Moreover, practitioners would be wise to assess the sensitivity of the estimators to various values of θ. We now illustrate this approach with an application.

4   Application

4.1   Background

To illustrate the estimation approach advocated herein using actual data, we revisit the analysis of school nutrition programs and childhood obesity in Millimet et al. (2008; hereafter MTH). As is quite evident from recent media reports, childhood obesity is deemed to have reached epidemic status (see Rosin (2008) for a review). In light of this, school nutrition programs, particularly the School Breakfast Program (SBP), have garnered much interest. We are interested in the role of this program in the obesity crisis.

Since extensive institutional details are provided in MTH, we simply provide a brief sketch of the SBP here. The SBP is federally funded, overseen by the US Department of Agriculture, administered by state education agencies, and been in existence for several decades. Participation by schools – both public and private – is voluntary (unless mandated by the state). If schools do participate, they are reimbursed a fixed amount per breakfast served. However, to qualify for reimbursement, the meals must meet federal nutrition guidelines. Finally, students residing in households with family incomes at or below 130% of the federal poverty line are eligible for free meals, while those in households with family incomes between 130% and 185% of the federal poverty line are entitled to reduced price meals. Schools are reimbursed at a higher rate per free or reduced price meals served. Currently, about 10 million students eat breakfast at school on any given day (covering about 80% of all schools).

Researchers interested in analyzing the causal effect of participation in either program have been concerned with the possible non-random selection of students into the program. MTH find evidence of positive selection on weight (in levels and growth rates) into the SBP (see also Bhattacharya et al. (2006)). In light of the Monte Carlo results above, this suggests a lower value of θ is more appropriate when analyzing the impact of participation in the SBP.

4.2 Data

The data are obtained from *Early Childhood Longitudinal Study-Kindergarten Class of 1998-99* (ECLS-K) and are identical to MTH; thus, we only provide limited details. We measure of participation in the SBP at the earliest possible date, which is in spring kindergarten. Our outcomes of interest are measures of child health in spring third grade or the change from fall kindergarten to spring third grade. As such, we are analyzing more of the long-run relationship between child health and participation in these two programs. Specifically, we utilize seven measures of child health:

  (i) body mass index (BMI) in levels in spring third grade,

  (ii) BMI in logs in spring third grade,

  (iii) growth rate in BMI (i.e., change in log BMI) from fall kindergarten to spring third grade,

  (iv) BMI in percentile in spring third grade,

  (v) change in BMI percentile from fall kindergarten to spring third grade,

  (vi) indicator for overweight status in spring third grade, and

 (vii) indicator for obesity status in spring third grade,

where we define overweight (obesity) as having a BMI above the (85th) 95th percentile.[5]

To control for student, parental, and environmental factors, the following covariates are included in X: child's race (white, black, Hispanic, Asian, and other) and gender, child's birthweight, household income, mother's employment status, mother's education, number of children's books at home, mother's age at first birth, an indicator if the child's mother received WIC benefits during pregnancy, region, city type (urban, suburban, or rural), and the amount of food in the household.[6]

Given the nature of our data, children with missing data for gender and race are dropped from our sample. Missing values for the remaining control variables are imputed and imputation dummies are added to the control set. The final sample contains 13,534 students, of which 3,161 participate in the SBP. Table 11 provides summary statistics.

4.3 Results

The results are presented in Table 12. We report OLS estimates of the ATE in addition to the estimators considered in the Monte Carlo study. Given the increase in sample size relative to the Monte Carlo study,

---

[5] Percentiles are obtained using the -*zanthro*- command in Stata.

[6] Except for maternal employment, all controls come from either the fall or spring kindergarten survey.

we also set $\theta = 0.01$ and 0.03. In addition, we report the coefficient estimates on the selection terms included in the BVN model, as well as the corresponding BMPS, P^*, and bias obtained from (12). Finally, all standard errors and 90% confidence intervals are obtained using 500 bootstrap repetitions.[7]

Turning to Table 12, three findings stand out. First, consonant with MTH, there is evidence of selection on unobservables. Specifically, $\widehat{\rho_{0u}\sigma_0}$ is positive and statistically significant in five of the seven models. While $\widehat{\rho_{\delta u}\sigma_\delta}$ is negative in six of the seven models, the estimates are never statistically significant. Thus, the evidence suggest that unobservables associated with higher weight in the untreated state are positively correlated with treatment participation. However, there is no statistically meaningful evidence that children select into the treatment on the basis of unobserved, individual-specific effects of treatment.

Second, the estimated ATE is positive and statistically significant across all outcomes when using $\widehat{\tau}_{HI}$ or $\widehat{\tau}_{OLS}$. However, in light of the BVN estimates discussed above, as well as the evidence provided in MTH, it is unlikely that these estimators are unbiased. The BVN estimates of the ATEs, on the other hand, are negative in five of seven models, although the estimates are very imprecise; the standard errors are between 13 and 17 times larger than those for $\widehat{\tau}_{HI}$.

Finally, our minimum biased estimators yield estimates that are always statistically insignificant. This arises for two reasons. First, given the evidence provided here and in MTH concerning selection on unobservables, the point estimates when $\theta = 0.05$, for example, are smaller than the corresponding Hirano and Imbens (2001) estimator in five of seven cases. Moving to $\theta = 0.01$, the point estimates become even smaller, and are negative, in four of seven cases. Second, the standard errors when $\theta = 0.05$ are roughly two to three times larger than for $\widehat{\tau}_{HI}$; the standard errors are roughly twice as large when $\theta = 0.01$ than when $\theta = 0.05$. Nonetheless, the fact that the point estimates diminish in the majority of cases as θ gets smaller, and four are negative for the smallest value of θ, is consistent with our expectations and illustrates the usefulness of our approach.

5   Conclusion

Black and Smith (2004) propose restricting the estimation sample to observations with propensity scores around one-half when estimating the ATT under unconfoundedness, but the researcher is worried that this assumption may not hold. In this study, we extend this argument to the case where the researcher may be interested in other average treatment effect parameters such as the ATU or ATE. This extension reveals two interesting findings. First, under non-negative selection on individual-specific, unobserved gains to treatment, the bias that results when unconfoundedness fails to hold is smallest when estimating the ATT.

---

[7]Confidence intervals are obtained using the percentile method. A school-level clustered bootstrap was also utilized; the confidence intervals always included zero (results available upon request).

Second, the value of the propensity score that minimizes the bias of estimators of the ATE is not fixed, but rather depends on the correlation structure of the unobservables in the underlying data-generating process.

To operationalize this knowledge, we have proposed a new, two-stage estimation technique of the ATE. In the first-stage, one estimates the usual Heckman bivariate normal selection model in order to recover estimates of the error correlation structure. These estimates are then used to obtain the value of the propensity score that minimizes the bias of selection on observables estimators of the ATE when unconfoundedness fails. In the second-stage, one estimates the ATE using only observations within a neighborhood around this bias-minimizing value of the propensity score.

Simulated data reveals that our approach is extremely useful in certain situations, and thus provides a nice addition to existing methods. Importantly, we show that our approach improves upon traditional selection on observables estimators not only when unconfoundedness does not hold and the usual Heckman bivariate normal selection model is mis-specified, but also when unconfoundedness holds. Future work should consider how θ may be chosen optimally in some sense, how observations may be differentially weighted depending on their proximity to the bias-minimizing propensity score, and how our method performs when the effect of treatment varies with observables.

References

[1] Bhattacharya, J., J. Currie, and S. Haider (2006), "Breakfast of Champions? The School Breakfast Program and the Nutrition of Children and Families," *Journal of Human Resources*, 41, 445-466.

[2] Black, D.A. and J.A. Smith (2004), "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching," *Journal of Econometrics*, 121, 99-124.

[3] D'Agostino, R.B., Jr. (1998), "Tutorial in Biostatistics: Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-randomized Control Group," *Statistics in Medicine*, 17, 2265-2281.

[4] Fisher, R.A. (1935), *The Design of Experiments*, Edinburgh: Oliver & Boyd.

[5] Heckman, J. and E. Vytlacil (2005), "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica*, 73, 669-738.

[6] Hirano, K. and Imbens, G.W. (2001), "Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization," *Health Services and Outcomes Research Methodology*, 2, 259-278.

[7] Hirano, K., G.W. Imbens, and G. Ridder (2003), "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score," *Econometrica*, 71, 1161-1189.

[8] Horvitz, D.G. and D.J. Thompson (1952), "A Generalization of Sampling Without Replacement from a Finite Universe," *Journal of the American Statistical Association*, 47, 663-685.

[9] Millimet, D.L. and R. Tchernis (2007), "On the Specification of Propensity Scores: with Applications to the Analysis of Trade Policies," *Journal of Business & Economic Statistics*, forthcoming.

[10] Millimet, D.L., R. Tchernis, and M. Hussain (2008), "School Nutrition Programs and the Incidence of Childhood Obesity," CAEPR Working Paper No. 2007-014.

[11] Neyman, J. (1923), "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," translated in *Statistical Science*, (with discussion), 5, 465-480, (1990).

[12] Rosin, O. (2008), "The Economic Causes of Obesity: A Survey," *Journal of Economic Surveys*, forthcoming.

[13] Roy, A.D. (1951), "Some Thoughts on the Distribution of Income," *Oxford Economic Papers*, 3, 135-146.

[14] Royston, P. (1991), "Comment on SG3.4 and an Improved D'Agostino Test," *Stata Technical Bulletin* 3:19. Reprinted in Stata Technical Bulletin Reprints, 1, 110-112.

[15] Rubin, D. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66, 688-701.

[16] Schanzenbach, D.W. (2007), "Does the Federal School Lunch Program Contribute to Childhood Obesity?" *Journal of Human Resources*, forthcoming.

[17] Smith, J.A. and P.E. Todd (2005), "Does Matching Overcome LaLonde's Critique?" *Journal of Econometrics*, 125, 305-353.
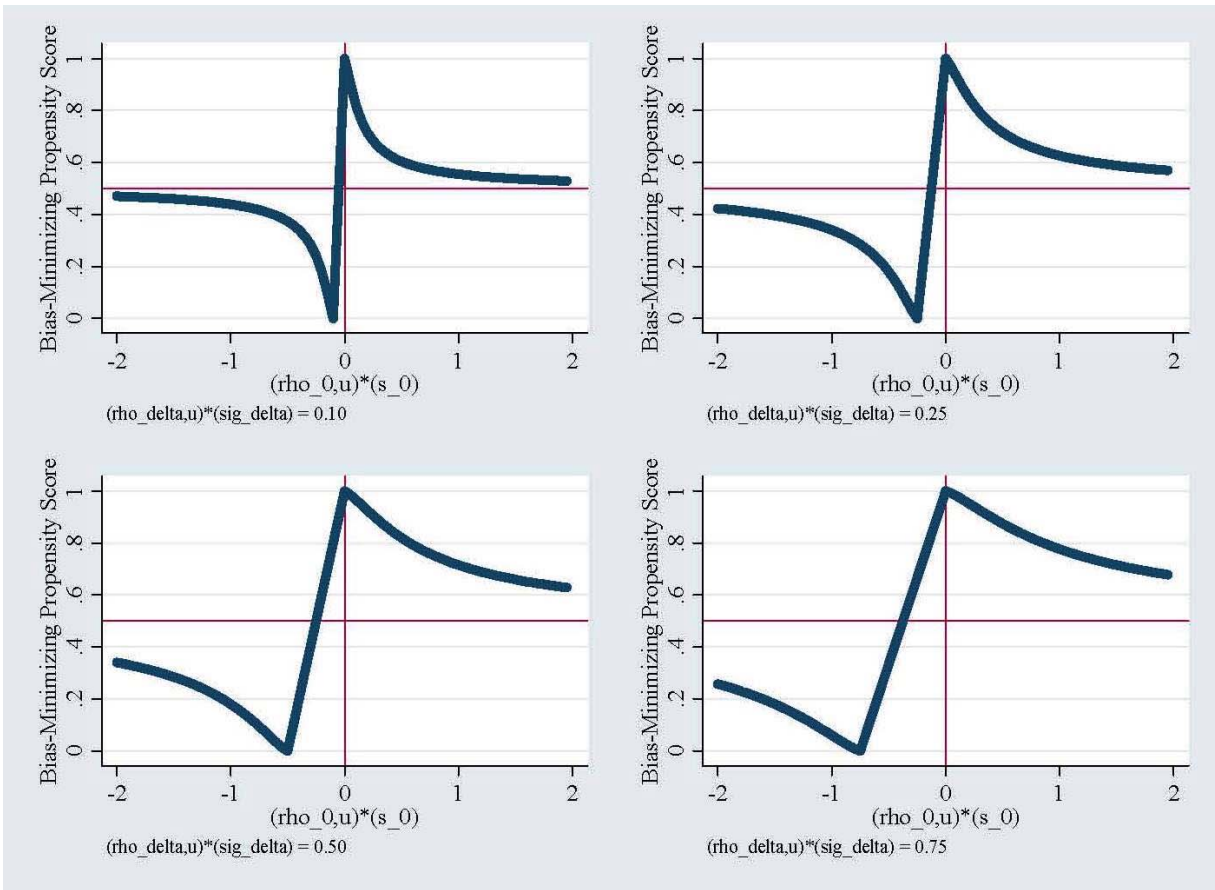
Figure 1. Bias-Minimizing Value of the Propensity Score Under Different Parameter Values Assuming Positive Selection on Unobserved, Individual-Specific Gains into Treatment.
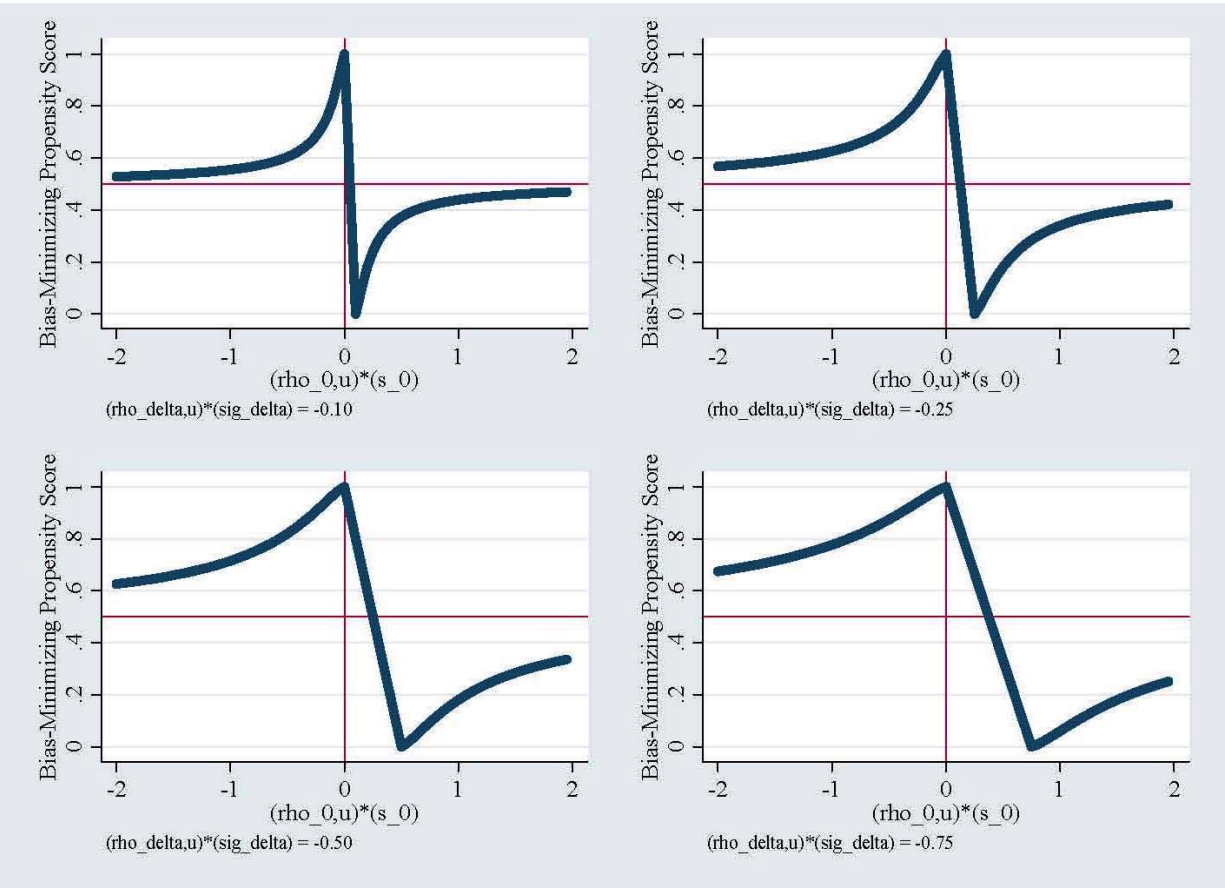
Figure 2. Bias-Minimizing Value of the Propensity Score Under Different Parameter Values Assuming Negative Selection on Unobserved, Individual-Specific Gains into Treatment.
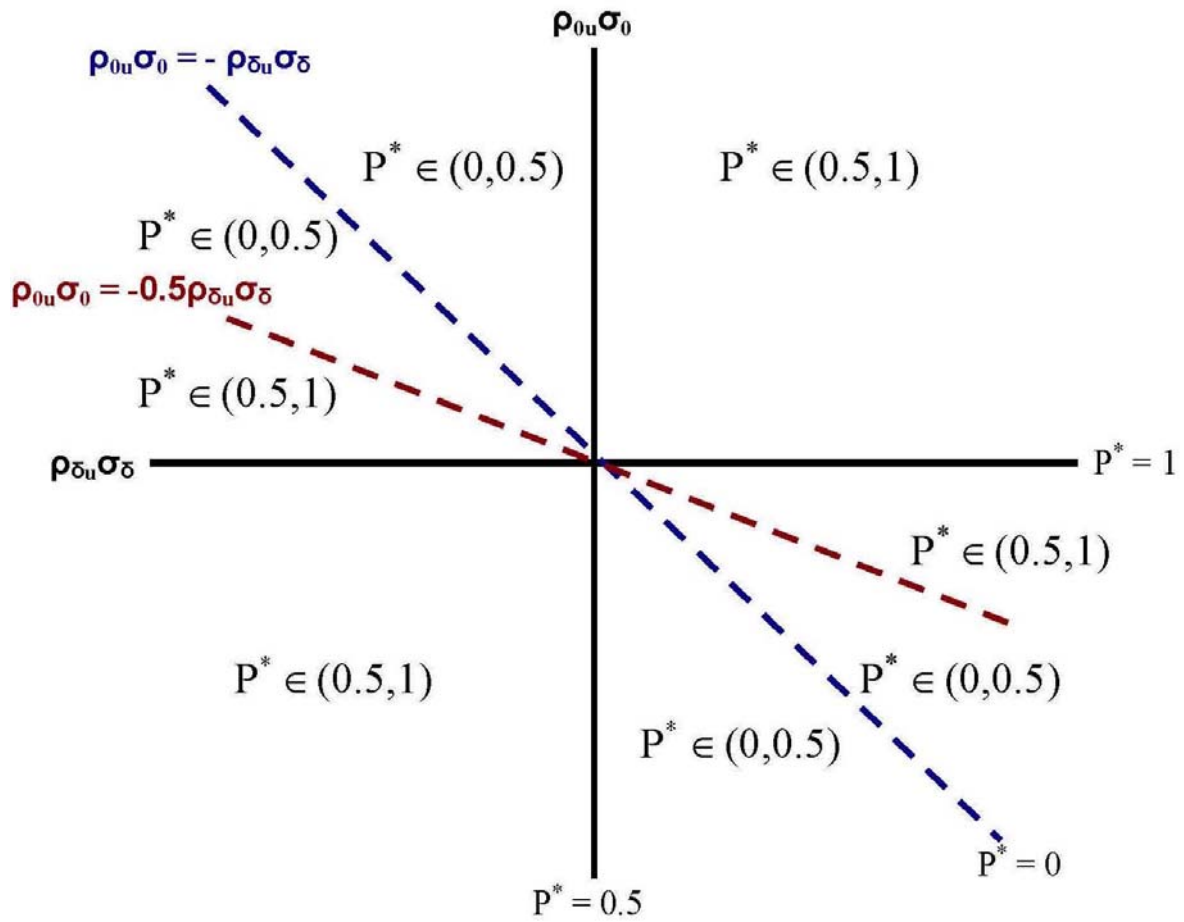
Figure 3. Bias-Minimizing Value of the Propensity Score Under Different Parameter Values.

Note: P^* denotes the bias-minimizing value of the propensity score.

Table 1. Monte Carlo Results: All Parametric Assumptions Valid.

	Constant Treatment Effect			Heterogeneous Treatment Effect								
True $\rho_0\sigma_0$ =	0	0.25	0.50	0			0.15			0.30		
True $\rho_\delta\sigma_\delta$ =				0	0.25	0.50	0	0.25	0.50	0	0.25	0.50
MEAN ESTIMATE												
$\rho_0\sigma_0$	-0.006	0.248	0.496	0.005	-0.015	-0.003	0.145	0.140	0.137	0.308	0.301	0.301
$\rho_\delta\sigma_\delta$				-0.004	0.262	0.503	0.004	0.258	0.510	-0.016	0.243	0.491
$\alpha_{0.25}$	0.326	0.324	0.326	0.559	0.587	0.660	0.580	0.603	0.695	0.500	0.558	0.652
$\alpha_{0.10}$	0.157	0.156	0.156	0.297	0.370	0.459	0.307	0.400	0.492	0.245	0.358	0.450
$\alpha_{0.05}$	0.088	0.089	0.089	0.184	0.237	0.306	0.188	0.257	0.333	0.142	0.220	0.293
BIAS												
τ_{BVN}	0.005	0.011	0.010	-0.001	0.005	0.003	0.004	-0.001	0.010	0.005	0.002	0.009
τ_{HI}	0.005	0.462	0.927	0.020	0.251	0.529	0.281	0.542	0.799	0.562	0.812	1.074
$\tau_{MB,0.25}$	-0.004	0.420	0.827	0.011	0.176	0.375	0.241	0.443	0.631	0.499	0.713	0.898
$\tau_{MB,0.10}$	-0.001	0.403	0.806	0.028	0.128	0.275	0.220	0.386	0.521	0.468	0.662	0.805
$\tau_{MB,0.05}$	-0.011	0.401	0.818	0.022	0.123	0.222	0.215	0.376	0.473	0.452	0.628	0.765
MAE												
τ_{BVN}	0.096	0.093	0.096	0.095	0.091	0.092	0.093	0.087	0.089	0.088	0.092	0.090
τ_{HI}	0.170	0.463	0.927	0.158	0.273	0.529	0.295	0.542	0.799	0.562	0.812	1.074
$\tau_{MB,0.25}$	0.115	0.420	0.827	0.146	0.220	0.379	0.256	0.445	0.631	0.500	0.713	0.898
$\tau_{MB,0.10}$	0.151	0.405	0.806	0.169	0.219	0.309	0.257	0.402	0.525	0.470	0.663	0.806
$\tau_{MB,0.05}$	0.206	0.415	0.818	0.209	0.253	0.299	0.278	0.409	0.491	0.458	0.629	0.767
MSE												
τ_{BVN}	0.015	0.014	0.014	0.014	0.013	0.013	0.014	0.012	0.012	0.012	0.013	0.013
τ_{HI}	0.045	0.244	0.884	0.038	0.101	0.312	0.114	0.327	0.669	0.343	0.686	1.178
$\tau_{MB,0.25}$	0.021	0.196	0.701	0.033	0.069	0.179	0.088	0.231	0.434	0.277	0.540	0.836
$\tau_{MB,0.10}$	0.036	0.200	0.682	0.045	0.073	0.136	0.093	0.206	0.334	0.256	0.486	0.700
$\tau_{MB,0.05}$	0.069	0.228	0.726	0.069	0.097	0.135	0.111	0.225	0.314	0.265	0.459	0.660

Notes: Results based on 500 simulated data sets, each containing 1000 observations. MAE = mean absolute bias; MSE = mean squared error. BVN = Heckman bivariate normal selection model; HI = Hirano and Imbens (2001) estimator; MB = minimum biased estimator using a cut-off level (α) chosen to retain 25%, 10%, or 5% of the treatment and control groups. Shaded results indicate the best performance. See text for further details.

Table 2. Monte Carlo Results: Functional Form Mis-Specified for Outcome Equation.

	Constant Treatment Effect			Heterogeneous Treatment Effect								
True $\rho_0\sigma_0 =$	0	0.25	0.50	0			0.15			0.30		
True $\rho_\delta\sigma_\delta =$				0	0.25	0.50	0	0.25	0.50	0	0.25	0.50
MEAN ESTIMATE												
$\rho_0\sigma_0$	-5.757	-5.513	-5.265	-6.223	-6.258	-6.313	-6.129	-6.109	-6.127	-5.957	-5.992	-5.948
$\rho_\delta\sigma_\delta$				0.844	1.118	1.456	0.924	1.127	1.413	0.874	1.172	1.408
$\alpha_{0.25}$	0.326	0.324	0.326	0.381	0.394	0.425	0.388	0.398	0.425	0.380	0.404	0.433
$\alpha_{0.10}$	0.157	0.156	0.156	0.142	0.137	0.139	0.142	0.138	0.138	0.139	0.138	0.140
$\alpha_{0.05}$	0.088	0.089	0.089	0.076	0.071	0.069	0.076	0.072	0.069	0.074	0.071	0.069
BIAS												
τ_{BVN}	10.042	10.064	10.065	10.030	10.046	10.065	10.043	10.042	10.044	10.054	10.062	10.007
τ_{HI}	0.005	0.462	0.927	0.020	0.251	0.529	0.281	0.542	0.799	0.562	0.812	1.074
$\tau_{MB,0.25}$	-0.004	0.419	0.827	-0.006	0.275	0.598	0.244	0.544	0.858	0.509	0.801	1.125
$\tau_{MB,0.10}$	-0.001	0.403	0.807	0.001	0.261	0.542	0.256	0.501	0.796	0.489	0.748	1.048
$\tau_{MB,0.05}$	-0.012	0.403	0.815	0.002	0.261	0.541	0.261	0.488	0.781	0.488	0.740	1.032
MAE												
τ_{BVN}	10.042	10.064	10.065	10.030	10.046	10.065	10.043	10.042	10.044	10.054	10.062	10.007
τ_{HI}	0.170	0.463	0.927	0.158	0.273	0.529	0.295	0.542	0.799	0.562	0.812	1.074
$\tau_{MB,0.25}$	0.115	0.420	0.827	0.130	0.280	0.598	0.256	0.544	0.858	0.509	0.801	1.125
$\tau_{MB,0.10}$	0.151	0.405	0.807	0.159	0.277	0.542	0.270	0.502	0.796	0.489	0.748	1.048
$\tau_{MB,0.05}$	0.205	0.416	0.815	0.221	0.296	0.544	0.299	0.492	0.781	0.491	0.740	1.032
MSE												
τ_{BVN}	101.017	101.452	101.475	100.782	101.119	101.474	101.029	101.012	101.051	101.255	101.418	100.302
τ_{HI}	0.045	0.244	0.884	0.038	0.101	0.312	0.114	0.327	0.669	0.343	0.686	1.178
$\tau_{MB,0.25}$	0.021	0.196	0.702	0.027	0.104	0.384	0.086	0.323	0.761	0.281	0.664	1.285
$\tau_{MB,0.10}$	0.037	0.200	0.684	0.039	0.105	0.327	0.098	0.285	0.667	0.270	0.594	1.122
$\tau_{MB,0.05}$	0.068	0.229	0.721	0.074	0.131	0.352	0.131	0.292	0.668	0.295	0.605	1.109

Notes: See Table 1.

Table 3. Monte Carlo Results: Functional Form Mis-Specified for Propensity Score Equation and Outcome Equation.

	Constant Treatment Effect			Heterogeneous Treatment Effect								
True $\rho_0\sigma_0 =$	0	0.25	0.50	0			0.15			0.30		
True $\rho_\delta\sigma_\delta =$				0	0.25	0.50	0	0.25	0.50	0	0.25	0.50
MEAN ESTIMATE												
$\rho_0\sigma_0$	85.170	84.426	82.129	123.228	119.039	118.633	121.819	120.524	119.603	120.478	119.015	116.419
$\rho_\delta\sigma_\delta$				-16.884	-16.443	-16.328	-16.907	-16.619	-16.482	-16.524	-16.700	-16.203
$\alpha_{0.25}$	0.073	0.074	0.073	0.121	0.122	0.121	0.122	0.121	0.121	0.121	0.122	0.123
$\alpha_{0.10}$	0.033	0.034	0.034	0.072	0.072	0.072	0.073	0.072	0.072	0.071	0.073	0.072
$\alpha_{0.05}$	0.020	0.020	0.020	0.045	0.045	0.044	0.045	0.045	0.045	0.044	0.045	0.045
BIAS												
τ_{BVN}	-132.212	-130.875	-127.001	-178.939	-172.468	-171.819	-176.506	-174.594	-173.124	-174.625	-171.991	-168.184
τ_{HI}	7.619	7.757	7.902	7.641	7.702	7.784	7.739	7.763	7.826	7.775	7.852	7.934
$\tau_{MB,0.25}$	3.649	3.892	4.103	3.235	3.429	3.605	3.405	3.561	3.773	3.519	3.759	3.929
$\tau_{MB,0.10}$	3.117	3.362	3.637	2.582	2.868	3.115	2.778	3.030	3.326	2.964	3.278	3.524
$\tau_{MB,0.05}$	2.983	3.265	3.520	2.404	2.717	3.019	2.617	2.905	3.250	2.818	3.183	3.440
MAE												
τ_{BVN}	132.212	130.875	127.001	178.939	172.468	171.819	176.506	174.594	173.124	174.625	171.991	168.184
τ_{HI}	7.619	7.757	7.902	7.641	7.702	7.784	7.739	7.763	7.826	7.775	7.852	7.934
$\tau_{MB,0.25}$	3.649	3.892	4.103	3.235	3.429	3.605	3.405	3.561	3.773	3.519	3.759	3.929
$\tau_{MB,0.10}$	3.117	3.362	3.637	2.582	2.868	3.115	2.778	3.030	3.326	2.964	3.278	3.524
$\tau_{MB,0.05}$	2.983	3.265	3.520	2.404	2.717	3.019	2.617	2.905	3.250	2.818	3.183	3.440
MSE												
τ_{BVN}	1.9E+04	1.9E+04	1.7E+04	3.4E+04	3.2E+04	3.2E+04	3.4E+04	3.3E+04	3.2E+04	3.2E+04	3.2E+04	3.0E+04
τ_{HI}	58.219	60.320	62.606	58.561	59.474	60.760	60.060	60.404	61.390	60.623	61.800	63.099
$\tau_{MB,0.25}$	13.521	15.380	17.054	10.586	11.890	13.127	11.705	12.811	14.358	12.495	14.231	15.545
$\tau_{MB,0.10}$	10.085	11.679	13.616	6.840	8.417	9.895	7.881	9.362	11.244	8.951	10.886	12.566
$\tau_{MB,0.05}$	9.421	11.232	12.959	6.000	7.628	9.355	7.062	8.677	10.824	8.156	10.324	12.020

Notes: See Table 1.

Table 4. Monte Carlo Results: Error Distribution is Mis-specified.

	Constant Treatment Effect			Heterogeneous Treatment Effect								
True $\rho_0\sigma_0 =$	**0**	**0.25**	**0.50**	**0**			**0.15**			**0.30**		
True $\rho_\delta\sigma_\delta =$				**0**	**0.25**	**0.50**	**0**	**0.25**	**0.50**	**0**	**0.25**	**0.50**
MEAN ESTIMATE												
$\rho_0\sigma_0$	-0.004	0.142	0.283	0.009	-0.007	0.006	0.084	0.078	0.079	0.172	0.172	0.176
$\rho_\delta\sigma_\delta$				-0.008	0.149	0.277	0.006	0.147	0.288	-0.009	0.136	0.272
$\alpha_{0.25}$	0.482	0.471	0.479	0.728	0.734	0.796	0.731	0.745	0.825	0.684	0.701	0.785
$\alpha_{0.10}$	0.249	0.249	0.248	0.401	0.474	0.564	0.412	0.502	0.597	0.367	0.462	0.555
$\alpha_{0.05}$	0.145	0.144	0.145	0.225	0.294	0.370	0.230	0.314	0.397	0.197	0.287	0.358
BIAS												
τ_{BVN}	0.002	0.008	0.010	-0.001	0.003	0.003	0.002	0.003	0.010	0.007	0.005	0.011
τ_{HI}	0.071	0.331	0.576	0.086	0.216	0.378	0.226	0.366	0.507	0.374	0.513	0.663
$\tau_{MB,0.25}$	-0.040	0.226	0.482	0.015	0.130	0.271	0.159	0.273	0.404	0.305	0.423	0.570
$\tau_{MB,0.10}$	-0.038	0.207	0.456	0.037	0.125	0.242	0.157	0.253	0.359	0.292	0.408	0.530
$\tau_{MB,0.05}$	-0.023	0.223	0.477	0.050	0.128	0.236	0.150	0.254	0.353	0.283	0.402	0.522
MAE												
τ_{BVN}	0.049	0.048	0.051	0.049	0.046	0.047	0.047	0.045	0.045	0.045	0.046	0.047
τ_{HI}	0.103	0.332	0.576	0.107	0.219	0.378	0.227	0.366	0.507	0.374	0.513	0.663
$\tau_{MB,0.25}$	0.102	0.228	0.482	0.082	0.138	0.271	0.165	0.273	0.404	0.305	0.423	0.570
$\tau_{MB,0.10}$	0.102	0.212	0.456	0.105	0.152	0.245	0.172	0.255	0.359	0.292	0.408	0.530
$\tau_{MB,0.05}$	0.127	0.232	0.477	0.132	0.165	0.247	0.177	0.261	0.355	0.285	0.403	0.522
MSE												
τ_{BVN}	0.004	0.004	0.004	0.004	0.003	0.004	0.004	0.003	0.003	0.003	0.003	0.003
τ_{HI}	0.016	0.118	0.339	0.017	0.057	0.151	0.061	0.142	0.265	0.148	0.271	0.446
$\tau_{MB,0.25}$	0.016	0.062	0.242	0.010	0.026	0.082	0.036	0.083	0.172	0.102	0.189	0.332
$\tau_{MB,0.10}$	0.016	0.057	0.219	0.017	0.033	0.075	0.041	0.081	0.146	0.098	0.182	0.295
$\tau_{MB,0.05}$	0.025	0.074	0.247	0.026	0.041	0.082	0.046	0.089	0.149	0.101	0.183	0.291

Notes: See Table 1.

Table 5. Monte Carlo Results: Functional Form Mis-Specified for Outcome Equation and Error Distribution.

	Constant Treatment Effect			Heterogeneous Treatment Effect								
True $\rho_0\sigma_0$ =	0	0.25	0.50	0			0.15			0.30		
True $\rho_\delta\sigma_\delta$ =				0	0.25	0.50	0	0.25	0.50	0	0.25	0.50
MEAN ESTIMATE												
$\rho_0\sigma_0$	-5.053	-4.915	-4.778	-5.576	-5.612	-5.658	-5.536	-5.520	-5.514	-5.448	-5.461	-5.439
$\rho_\delta\sigma_\delta$				0.907	1.097	1.303	0.981	1.095	1.223	0.975	1.130	1.255
$\alpha_{0.25}$	0.482	0.471	0.479	0.571	0.585	0.610	0.587	0.593	0.602	0.581	0.592	0.612
$\alpha_{0.10}$	0.249	0.249	0.248	0.230	0.228	0.232	0.234	0.230	0.229	0.230	0.231	0.231
$\alpha_{0.05}$	0.145	0.144	0.145	0.126	0.123	0.121	0.128	0.124	0.120	0.127	0.124	0.119
BIAS												
τ_{BVN}	8.886	8.905	8.915	8.896	8.899	8.920	8.900	8.885	8.898	8.893	8.908	8.888
τ_{HI}	0.071	0.331	0.576	0.086	0.216	0.378	0.226	0.366	0.507	0.374	0.513	0.663
$\tau_{MB,0.25}$	-0.040	0.227	0.483	0.030	0.173	0.342	0.182	0.321	0.479	0.325	0.476	0.632
$\tau_{MB,0.10}$	-0.039	0.207	0.456	-0.028	0.138	0.317	0.124	0.281	0.466	0.267	0.438	0.617
$\tau_{MB,0.05}$	-0.023	0.224	0.477	-0.010	0.135	0.293	0.142	0.269	0.440	0.270	0.424	0.597
MAE												
τ_{BVN}	8.886	8.905	8.915	8.896	8.899	8.920	8.900	8.885	8.898	8.893	8.908	8.888
τ_{HI}	0.103	0.332	0.576	0.107	0.219	0.378	0.227	0.366	0.507	0.374	0.513	0.663
$\tau_{MB,0.25}$	0.102	0.229	0.483	0.083	0.177	0.342	0.186	0.321	0.479	0.326	0.476	0.632
$\tau_{MB,0.10}$	0.102	0.211	0.456	0.103	0.156	0.317	0.143	0.282	0.466	0.267	0.438	0.617
$\tau_{MB,0.05}$	0.126	0.233	0.477	0.131	0.168	0.297	0.171	0.274	0.440	0.273	0.424	0.597
MSE												
τ_{BVN}	79.095	79.429	79.615	79.271	79.331	79.700	79.346	79.072	79.288	79.216	79.482	79.134
τ_{HI}	0.016	0.118	0.339	0.017	0.057	0.151	0.061	0.142	0.265	0.148	0.271	0.446
$\tau_{MB,0.25}$	0.016	0.063	0.242	0.011	0.040	0.125	0.043	0.112	0.238	0.115	0.235	0.405
$\tau_{MB,0.10}$	0.016	0.057	0.219	0.016	0.035	0.114	0.028	0.093	0.229	0.084	0.204	0.389
$\tau_{MB,0.05}$	0.025	0.074	0.247	0.026	0.043	0.109	0.043	0.096	0.212	0.092	0.199	0.369

Notes: See Table 1.

Table 6. Monte Carlo Results: Comparison of Using True P^* vs. Estimated P* When All Parametric Assumptions are Valid.

	Constant Treatment Effect			Heterogeneous Treatment Effect								
True $\rho_0\sigma_0 =$	0	0.25	0.50	0			0.15			0.30		
True $\rho_\delta\sigma_\delta =$				0	0.25	0.50	0	0.25	0.50	0	0.25	0.50
MEAN												
Estimated P^*	0.499	0.499	0.499	0.508	0.730	0.876	0.491	0.803	0.912	0.482	0.763	0.866
True P^*		0.498	0.498		0.980	0.980	0.498	0.886	0.947	0.498	0.793	0.886
BIAS (Estimated P)*												
$\tau_{MB,0.25}$	-0.004	0.420	0.827	0.011	0.176	0.375	0.241	0.443	0.631	0.499	0.713	0.898
$\tau_{MB,0.10}$	-0.001	0.403	0.806	0.028	0.128	0.275	0.220	0.386	0.521	0.468	0.662	0.805
$\tau_{MB,0.05}$	-0.011	0.401	0.818	0.022	0.123	0.222	0.215	0.376	0.473	0.452	0.628	0.765
BIAS (True P)*												
$\tau_{MB,0.25}$		0.419	0.827		0.171	0.374	0.250	0.447	0.630	0.489	0.720	0.901
$\tau_{MB,0.10}$		0.403	0.807		0.115	0.269	0.248	0.400	0.522	0.488	0.686	0.809
$\tau_{MB,0.05}$		0.402	0.817		0.094	0.210	0.235	0.394	0.477	0.478	0.678	0.779
MAE (Estimated P)*												
$\tau_{MB,0.25}$	0.115	0.420	0.827	0.146	0.220	0.379	0.256	0.445	0.631	0.500	0.713	0.898
$\tau_{MB,0.10}$	0.151	0.405	0.806	0.169	0.219	0.309	0.257	0.402	0.525	0.470	0.663	0.806
$\tau_{MB,0.05}$	0.206	0.415	0.818	0.209	0.253	0.299	0.278	0.409	0.491	0.458	0.629	0.767
MAE (True P)*												
$\tau_{MB,0.25}$		0.420	0.827		0.227	0.379	0.255	0.450	0.630	0.489	0.720	0.901
$\tau_{MB,0.10}$		0.405	0.807		0.241	0.311	0.266	0.418	0.526	0.488	0.687	0.809
$\tau_{MB,0.05}$		0.415	0.817		0.274	0.302	0.282	0.433	0.495	0.483	0.679	0.781
MSE (Estimated P)*												
$\tau_{MB,0.25}$	0.021	0.196	0.701	0.033	0.069	0.179	0.088	0.231	0.434	0.277	0.540	0.836
$\tau_{MB,0.10}$	0.036	0.200	0.682	0.045	0.073	0.136	0.093	0.206	0.334	0.256	0.486	0.700
$\tau_{MB,0.05}$	0.069	0.228	0.726	0.069	0.097	0.135	0.111	0.225	0.314	0.265	0.459	0.660
MSE (True P)*												
$\tau_{MB,0.25}$		0.196	0.702		0.073	0.178	0.083	0.238	0.434	0.258	0.551	0.841
$\tau_{MB,0.10}$		0.199	0.685		0.086	0.137	0.097	0.225	0.336	0.273	0.526	0.708
$\tau_{MB,0.05}$		0.229	0.727		0.111	0.137	0.123	0.248	0.321	0.286	0.535	0.691

Notes: Results using estimated correlation structure are replicated from Table 1. Results using true P* utilize the P* corresponding to the true values of $\rho_0\sigma_0$ and $\rho_\delta\sigma_\delta$. Shaded area indicates best overall performance. True P* results are omitted under unconfoundedness since there is no unique value of P*. For further details, see Table 1.

Table 7. Monte Carlo Results: Comparison of Using True P* vs. Estimated P* When the Functional Form Is Mis-Specified for the Outcome Equation.

	Constant Treatment Effect			Heterogeneous Treatment Effect								
True $\rho_0\sigma_0 =$	**0**	**0.25**	**0.50**	**0**			**0.15**			**0.30**		
True $\rho_\delta\sigma_\delta =$				**0**	**0.25**	**0.50**	**0**	**0.25**	**0.50**	**0**	**0.25**	**0.50**
MEAN												
Estimated P*	0.499	0.499	0.499	0.418	0.390	0.356	0.408	0.387	0.356	0.411	0.379	0.352
True P*		0.498	0.498		0.980	0.980	0.498	0.886	0.947	0.498	0.793	0.886
BIAS (Estimated P*)												
$\tau_{MB,0.25}$	-0.004	0.419	0.827	-0.006	0.275	0.598	0.244	0.544	0.858	0.509	0.801	1.125
$\tau_{MB,0.10}$	-0.001	0.403	0.807	0.001	0.261	0.542	0.256	0.501	0.796	0.489	0.748	1.048
$\tau_{MB,0.05}$	-0.012	0.403	0.815	0.002	0.261	0.541	0.261	0.488	0.781	0.488	0.740	1.032
BIAS (True P*)												
$\tau_{MB,0.25}$		0.419	0.827		0.171	0.374	0.250	0.447	0.630	0.489	0.720	0.901
$\tau_{MB,0.10}$		0.403	0.807		0.115	0.269	0.248	0.400	0.522	0.488	0.686	0.809
$\tau_{MB,0.05}$		0.402	0.817		0.094	0.210	0.235	0.394	0.477	0.478	0.678	0.779
MAE (Estimated P*)												
$\tau_{MB,0.25}$	0.115	0.420	0.827	0.130	0.280	0.598	0.256	0.544	0.858	0.509	0.801	1.125
$\tau_{MB,0.10}$	0.151	0.405	0.807	0.159	0.277	0.542	0.270	0.502	0.796	0.489	0.748	1.048
$\tau_{MB,0.05}$	0.205	0.416	0.815	0.221	0.296	0.544	0.299	0.492	0.781	0.491	0.740	1.032
MAE (True P*)												
$\tau_{MB,0.25}$		0.420	0.827		0.227	0.379	0.255	0.450	0.630	0.489	0.720	0.901
$\tau_{MB,0.10}$		0.405	0.807		0.241	0.311	0.266	0.418	0.526	0.488	0.687	0.809
$\tau_{MB,0.05}$		0.415	0.817		0.274	0.302	0.282	0.433	0.495	0.483	0.679	0.781
MSE (Estimated P*)												
$\tau_{MB,0.25}$	0.021	0.196	0.702	0.027	0.104	0.384	0.086	0.323	0.761	0.281	0.664	1.285
$\tau_{MB,0.10}$	0.037	0.200	0.684	0.039	0.105	0.327	0.098	0.285	0.667	0.270	0.594	1.122
$\tau_{MB,0.05}$	0.068	0.229	0.721	0.074	0.131	0.352	0.131	0.292	0.668	0.295	0.605	1.109
MSE (True P*)												
$\tau_{MB,0.25}$		0.196	0.702		0.073	0.178	0.083	0.238	0.434	0.258	0.551	0.841
$\tau_{MB,0.10}$		0.199	0.685		0.086	0.137	0.097	0.225	0.336	0.273	0.526	0.708
$\tau_{MB,0.05}$		0.229	0.727		0.111	0.137	0.123	0.248	0.321	0.286	0.535	0.691

Notes: Results using estimated correlation structure are replicated from Table 2. For further details, see Table 6.

Table 8. Monte Carlo Results: Comparison of Using True P^* vs. Estimated P* When the Functional Form is Mis-Specified for the Propensity Score Equation and Outcome Equation.

	Constant Treatment Effect			Heterogeneous Treatment Effect								
True $\rho_0\sigma_0=$	0	0.25	0.50	0			0.15			0.30		
True $\rho_\delta\sigma_\delta=$				0	0.25	0.50	0	0.25	0.50	0	0.25	0.50
MEAN												
Estimated P^*	0.499	0.499	0.499	0.413	0.412	0.413	0.411	0.413	0.413	0.413	0.411	0.412
True P^*		0.498	0.498		0.980	0.980	0.498	0.886	0.947	0.498	0.793	0.886
BIAS (Estimated P*)												
$\tau_{MB,0.25}$	3.649	3.892	4.103	3.235	3.429	3.605	3.405	3.561	3.773	3.519	3.759	3.929
$\tau_{MB,0.10}$	3.117	3.362	3.637	2.582	2.868	3.115	2.778	3.030	3.326	2.964	3.278	3.524
$\tau_{MB,0.05}$	2.983	3.265	3.520	2.404	2.717	3.019	2.617	2.905	3.250	2.818	3.183	3.440
BIAS (True P*)												
$\tau_{MB,0.25}$		3.881	4.095		8.512	8.591	3.801	8.584	8.635	3.896	8.654	8.696
$\tau_{MB,0.10}$		3.348	3.621		11.111	11.182	3.286	11.176	11.183	3.422	11.165	11.162
$\tau_{MB,0.05}$		3.250	3.510		13.573	13.667	3.182	13.627	13.595	3.300	13.582	13.546
MAE (Estimated P*)												
$\tau_{MB,0.25}$	3.649	3.892	4.103	3.235	3.429	3.605	3.405	3.561	3.773	3.519	3.759	3.929
$\tau_{MB,0.10}$	3.117	3.362	3.637	2.582	2.868	3.115	2.778	3.030	3.326	2.964	3.278	3.524
$\tau_{MB,0.05}$	2.983	3.265	3.520	2.404	2.717	3.019	2.617	2.905	3.250	2.818	3.183	3.440
MAE (True P*)												
$\tau_{MB,0.25}$		3.881	4.095		8.512	8.591	3.801	8.584	8.635	3.896	8.654	8.696
$\tau_{MB,0.10}$		3.348	3.621		11.111	11.182	3.286	11.176	11.183	3.422	11.165	11.162
$\tau_{MB,0.05}$		3.250	3.510		13.573	13.667	3.182	13.627	13.595	3.300	13.582	13.546
MSE (Estimated P*)												
$\tau_{MB,0.25}$	13.521	15.380	17.054	10.586	11.890	13.127	11.705	12.811	14.358	12.495	14.231	15.545
$\tau_{MB,0.10}$	10.085	11.679	13.616	6.840	8.417	9.895	7.881	9.362	11.244	8.951	10.886	12.566
$\tau_{MB,0.05}$	9.421	11.232	12.959	6.000	7.628	9.355	7.062	8.677	10.824	8.156	10.324	12.020
MSE (True P*)												
$\tau_{MB,0.25}$		15.289	16.988		72.714	74.101	14.669	73.935	74.819	15.388	75.142	75.857
$\tau_{MB,0.10}$		11.580	13.501		124.571	126.405	11.201	126.023	126.321	12.098	125.815	125.639
$\tau_{MB,0.05}$		11.122	12.892		186.986	189.832	10.703	188.494	187.667	11.398	187.006	185.870

Notes: Results using estimated correlation structure are replicated from Table 3. For further details, see Table 6.

Table 9. Monte Carlo Results: Comparison of Using True P* vs. Estimated P* When the Error Distribution Is Mis-Specified.

	Constant Treatment Effect			Heterogeneous Treatment Effect								
True $\rho_0\sigma_0 =$	**0**	**0.25**	**0.50**	**0**			**0.15**			**0.30**		
True $\rho_\delta\sigma_\delta =$				**0**	**0.25**	**0.50**	**0**	**0.25**	**0.50**	**0**	**0.25**	**0.50**
MEAN												
Estimated P*	0.499	0.499	0.499	0.475	0.718	0.867	0.498	0.779	0.899	0.480	0.749	0.856
True P*		0.498	0.498		0.980	0.980	0.498	0.886	0.947	0.498	0.793	0.886
BIAS **(Estimated P*)**												
$\tau_{MB,0.25}$	-0.040	0.226	0.482	0.015	0.130	0.271	0.159	0.273	0.404	0.305	0.423	0.570
$\tau_{MB,0.10}$	-0.038	0.207	0.456	0.037	0.125	0.242	0.157	0.253	0.359	0.292	0.408	0.530
$\tau_{MB,0.05}$	-0.023	0.223	0.477	0.050	0.128	0.236	0.150	0.254	0.353	0.283	0.402	0.522
BIAS **(True P*)**												
$\tau_{MB,0.25}$		0.228	0.484		0.115	0.269	0.131	0.271	0.404	0.275	0.422	0.570
$\tau_{MB,0.10}$		0.207	0.456		0.127	0.243	0.109	0.272	0.360	0.258	0.428	0.534
$\tau_{MB,0.05}$		0.224	0.477		0.151	0.243	0.130	0.303	0.359	0.268	0.449	0.534
MAE **(Estimated P*)**												
$\tau_{MB,0.25}$	0.102	0.228	0.482	0.082	0.138	0.271	0.165	0.273	0.404	0.305	0.423	0.570
$\tau_{MB,0.10}$	0.102	0.212	0.456	0.105	0.152	0.245	0.172	0.255	0.359	0.292	0.408	0.530
$\tau_{MB,0.05}$	0.127	0.232	0.477	0.132	0.165	0.247	0.177	0.261	0.355	0.285	0.403	0.522
MAE **(True P*)**												
$\tau_{MB,0.25}$		0.230	0.484		0.133	0.270	0.145	0.271	0.404	0.275	0.422	0.570
$\tau_{MB,0.10}$		0.212	0.456		0.159	0.249	0.136	0.274	0.361	0.260	0.429	0.534
$\tau_{MB,0.05}$		0.233	0.477		0.188	0.255	0.160	0.307	0.361	0.272	0.450	0.534
MSE **(Estimated P*)**												
$\tau_{MB,0.25}$	0.016	0.062	0.242	0.010	0.026	0.082	0.036	0.083	0.172	0.102	0.189	0.332
$\tau_{MB,0.10}$	0.016	0.057	0.219	0.017	0.033	0.075	0.041	0.081	0.146	0.098	0.182	0.295
$\tau_{MB,0.05}$	0.025	0.074	0.247	0.026	0.041	0.082	0.046	0.089	0.149	0.101	0.183	0.291
MSE **(True P*)**												
$\tau_{MB,0.25}$		0.063	0.243		0.025	0.082	0.030	0.083	0.172	0.087	0.187	0.332
$\tau_{MB,0.10}$		0.057	0.219		0.037	0.076	0.027	0.093	0.147	0.080	0.201	0.300
$\tau_{MB,0.05}$		0.074	0.247		0.050	0.085	0.040	0.119	0.155	0.096	0.226	0.307

Notes: Results using estimated correlation structure are replicated from Table 4. For further details, see Table 6.

Table 10. Monte Carlo Results: Comparison of Using True P* vs. Estimated P* When the Functional Form Is Mis-Specified for the Outcome Equation and Error Distribution.

	Constant Treatment Effect			Heterogeneous Treatment Effect								
True $\rho_0\sigma_0 =$	0	0.25	0.50	0			0.15			0.30		
True $\rho_\delta\sigma_\delta =$				0	0.25	0.50	0	0.25	0.50	0	0.25	0.50
MEAN												
Estimated P*	0.499	0.498	0.498	0.401	0.379	0.357	0.391	0.378	0.362	0.390	0.372	0.356
True P*		0.498	0.498		0.980	0.980	0.498	0.886	0.947	0.498	0.793	0.886
BIAS (Estimated P*)												
$\tau_{MB,0.25}$	-0.040	0.227	0.483	0.030	0.173	0.342	0.182	0.321	0.479	0.325	0.476	0.632
$\tau_{MB,0.10}$	-0.039	0.207	0.456	-0.028	0.138	0.317	0.124	0.281	0.466	0.267	0.438	0.617
$\tau_{MB,0.05}$	-0.023	0.224	0.477	-0.010	0.135	0.293	0.142	0.269	0.440	0.270	0.424	0.597
BIAS (True P*)												
$\tau_{MB,0.25}$		0.228	0.484		0.115	0.269	0.131	0.271	0.404	0.275	0.422	0.570
$\tau_{MB,0.10}$		0.207	0.456		0.127	0.243	0.109	0.272	0.360	0.258	0.428	0.534
$\tau_{MB,0.05}$		0.224	0.477		0.151	0.243	0.130	0.303	0.359	0.268	0.449	0.534
MAE (Estimated P*)												
$\tau_{MB,0.25}$	0.102	0.229	0.483	0.083	0.177	0.342	0.186	0.321	0.479	0.326	0.476	0.632
$\tau_{MB,0.10}$	0.102	0.211	0.456	0.103	0.156	0.317	0.143	0.282	0.466	0.267	0.438	0.617
$\tau_{MB,0.05}$	0.126	0.233	0.477	0.131	0.168	0.297	0.171	0.274	0.440	0.273	0.424	0.597
MAE (True P*)												
$\tau_{MB,0.25}$		0.230	0.484		0.133	0.270	0.145	0.271	0.404	0.275	0.422	0.570
$\tau_{MB,0.10}$		0.212	0.456		0.159	0.249	0.136	0.274	0.361	0.260	0.429	0.534
$\tau_{MB,0.05}$		0.233	0.477		0.188	0.255	0.160	0.307	0.361	0.272	0.450	0.534
MSE (Estimated P*)												
$\tau_{MB,0.25}$	0.016	0.063	0.242	0.011	0.040	0.125	0.043	0.112	0.238	0.115	0.235	0.405
$\tau_{MB,0.10}$	0.016	0.057	0.219	0.016	0.035	0.114	0.028	0.093	0.229	0.084	0.204	0.389
$\tau_{MB,0.05}$	0.025	0.074	0.247	0.026	0.043	0.109	0.043	0.096	0.212	0.092	0.199	0.369
MSE (True P*)												
$\tau_{MB,0.25}$		0.063	0.243		0.025	0.082	0.030	0.083	0.172	0.087	0.187	0.332
$\tau_{MB,0.10}$		0.057	0.219		0.037	0.076	0.027	0.093	0.147	0.080	0.201	0.300
$\tau_{MB,0.05}$		0.074	0.247		0.050	0.085	0.040	0.119	0.155	0.096	0.226	0.307

Notes: Results using estimated correlation structure are replicated from Table 5. For further details, see Table 6.

Table 11. Summary Statistics

Variable	Full Sample Mean	Full Sample SD	SBP Participants Mean	SBP Participants SD	SBP Non-Participants Mean	SBP Non-Participants SD
SBP Participation (1 = Yes)	0.234	0.423	1	0	0	0
Third Grade Child Weight						
BMI	18.404	3.861	18.956	4.296	18.236	3.702
BMI Growth Rate	0.112	0.126	0.128	0.137	0.107	0.122
BMI percentile	62.326	30.105	65.662	29.794	61.309	30.128
Change in BMI Percentile	1.295	22.887	2.894	23.108	0.808	22.798
Overweight (1 = Yes)	0.325	0.468	0.368	0.482	0.312	0.463
Obese (1 = Yes)	0.171	0.377	0.208	0.406	0.160	0.367
Controls						
Age (in months)	110.767	4.356	110.869	4.389	110.737	4.346
Gender (1 = boy)	0.507	0.500	0.523	0.500	0.503	0.500
White (1 = Yes)	0.579	0.494	0.322	0.467	0.657	0.475
Black (1 = Yes)	0.138	0.345	0.312	0.463	0.085	0.279
Hispanic (1 = Yes)	0.174	0.379	0.239	0.427	0.154	0.361
Asian (1 = Yes)	0.054	0.226	0.042	0.200	0.057	0.233
Child's Birthweight (ounces)	118.284	20.040	115.806	21.452	119.039	19.529
Child's Birthweight (1 = Missing)	0.121	0.326	0.164	0.371	0.107	0.309
Central City (1 = Yes)	0.395	0.489	0.416	0.493	0.389	0.488
Urban Fringe & Large Town (1 = Yes)	0.377	0.485	0.259	0.438	0.413	0.492
Northeast (1 = Yes)	0.182	0.386	0.115	0.319	0.202	0.402
Midwest (1 = Yes)	0.250	0.433	0.194	0.395	0.268	0.443
South (1 = Yes)	0.346	0.476	0.508	0.500	0.297	0.457
Mother's Age at First Birth ≤ 19 Years Old (1 = Yes)	0.227	0.419	0.404	0.491	0.173	0.378
Mother's Age at First Birth is 20-29 Years Old (1 = Yes)	0.522	0.500	0.410	0.492	0.556	0.497
Mother's Age at First Birth (1 = Missing)	0.104	0.305	0.140	0.347	0.093	0.290
WIC Benefits During Pregnancy (1 = Yes)	0.339	0.473	0.620	0.485	0.253	0.435
WIC Benefits During Pregnancy (1 = Missing)	0.112	0.315	0.140	0.347	0.103	0.304
Mother's Education = High School (1 = Yes)	0.198	0.398	0.243	0.429	0.184	0.387
Mother's Education = Some College (1 = Yes)	0.281	0.450	0.227	0.419	0.298	0.457
Mother's Education = Bachelor's Degree (1 = Yes)	0.144	0.351	0.040	0.196	0.176	0.381
Mother's Education = Advanced College Degree (1 = Yes)	0.084	0.277	0.024	0.152	0.102	0.303
Mother's Education (1 = Missing)	0.209	0.407	0.285	0.452	0.186	0.389
Household Income (dollars)	52150	32034	34535	21588	57518	32765
Mother Employed During 3rd Grade (1 = Yes)	0.572	0.495	0.467	0.499	0.604	0.489
Mother Employed During 3rd Grade (1 = No)	0.204	0.403	0.230	0.421	0.197	0.397
Sufficient Food of Type Desired in Household (1 = Yes)	0.847	0.360	0.736	0.441	0.881	0.324
Sufficient Food, but not of Type Desired in Household (1 = Yes)	0.138	0.345	0.229	0.420	0.111	0.314
Sufficient Food (1 = Missing)	0.001	0.028	0.002	0.040	0.001	0.024
NSLP Participation (1 = Yes)	0.575	0.494	0.894	0.308	0.477	0.500
Number of Children's Books in Household	74.930	57.030	48.638	46.803	82.942	57.457
Number of Children's Books in Household (1 = Missing)	0.097	0.296	0.119	0.323	0.091	0.287

Notes: N = 13,534 (full sample); 3,161 (SBP participants); 10,373 (SBP non-participants). Data are from spring third grade wave of ECLS-K. Change in BMI percentile and BMI growth rate calculated using baseline data from fall kindergarten. Omitted category for race is 'other', city type is 'small town & rural', mother's age at first birth is greater than 29 years old, mother's employment is 'missing', mother's education is 'less than high school', and sufficient food is 'sometimes or often there is not enough to eat'.

Table 12. Estimates of Effect of SBP Participation

	BMI	ln(BMI)	BMI Growth	Percentile BMI	Change in Percentile BMI	Pr(Overweight)	Pr(Obese)
τ_{OLS}	0.645	0.031	0.015	3.649	1.618	0.051	0.056
	(0.149)	(0.007)	(0.004)	(0.984)	(0.799)	(0.016)	(0.014)
	[0.400, 0.900]	[0.019, 0.042]	[0.008, 0.023]	[1.910, 5.245]	[0.375, 2.972]	[0.024, 0.077]	[0.034, 0.079]
τ_{BVN}	1.425	0.041	-0.022	-2.472	-9.021	-0.053	-0.088
	(2.080)	(0.104)	(0.061)	(14.002)	(10.603)	(0.235)	(0.202)
	[-2.007, 4.351]	[-0.142, 0.205]	[-0.117, 0.085]	[-25.935, 19.667]	[-27.182, 7.817]	[-0.434, 0.327]	[-0.416, 0.237]
τ_{HI}	0.392	0.019	0.013	2.574	2.016	0.033	0.029
	(0.131)	(0.006)	(0.004)	(1.008)	(0.842)	(0.016)	(0.013)
	[0.161, 0.596]	[0.008, 0.030]	[0.006, 0.019]	[0.849, 4.148]	[0.436, 3.279]	[0.006, 0.057]	[0.008, 0.052]
$\tau_{MB,0.25}$	0.288	0.019	0.010	2.694	0.603	0.063	0.008
	(0.165)	(0.008)	(0.005)	(1.175)	(0.775)	(0.023)	(0.020)
	[0.075, 0.594]	[0.003, 0.031]	[0.002, 0.019]	[0.509, 4.276]	[-0.446, 1.990]	[-0.009, 0.068]	[-0.013, 0.055]
$\tau_{MB,0.10}$	-0.007	0.025	0.003	2.887	0.873	0.074	-0.005
	(0.239)	(0.012)	(0.007)	(1.746)	(1.131)	(0.031)	(0.028)
	[-0.098, 0.725]	[-0.004, 0.036]	[-0.001, 0.023]	[-0.567, 5.053]	[-0.907, 2.654]	[-0.025, 0.076]	[-0.023, 0.069]
$\tau_{MB,0.05}$	-0.017	0.013	-0.007	3.657	1.474	0.061	-0.007
	(0.331)	(0.017)	(0.010)	(2.554)	(1.589)	(0.041)	(0.033)
	[-0.212, 0.929]	[-0.012, 0.047]	[-0.005, 0.028]	[-1.918, 6.832]	[-1.375, 3.616]	[-0.037, 0.093]	[-0.032, 0.078]
$\tau_{MB,0.03}$	0.138	0.014	-0.013	2.654	0.532	0.078	-0.029
	(0.446)	(0.022)	(0.013)	(2.843)	(1.962)	(0.047)	(0.045)
	[-0.368, 1.099]	[-0.021, 0.053]	[-0.008, 0.033]	[-2.123, 6.800]	[-1.992, 4.027]	[-0.056, 0.097]	[-0.052, 0.098]
$\tau_{MB,0.01}$	-0.413	-0.024	0.014	7.601	-1.996	0.113	-0.104
	(0.724)	(0.035)	(0.021)	(5.064)	(3.619)	(0.074)	(0.064)
	[-0.921, 1.516]	[-0.038, 0.079]	[-0.018, 0.050]	[-5.504, 10.067]	[-5.896, 5.839]	[-0.095, 0.150]	[-0.076, 0.128]
$\rho_0\sigma_0$	0.901	0.044	0.021	5.367	0.771	0.084	0.079
	(0.380)	(0.018)	(0.013)	(3.171)	(2.108)	(0.048)	(0.036)
	[0.303, 1.536]	[0.014, 0.075]	[-0.000, 0.043]	[0.418, 10.691]	[-2.699, 4.054]	[0.012, 0.164]	[0.017, 0.137]
$\rho_\delta\sigma_\delta$	-1.604	-0.061	-0.003	-2.759	5.765	-0.038	-0.006
	(1.338)	(0.069)	(0.041)	(9.784)	(7.128)	(0.156)	(0.131)
	[-3.646, 0.772]	[-0.168, 0.058]	[-0.075, 0.064]	[-18.844, 14.201]	[-5.756, 17.586]	[-0.281, 0.222]	[-0.209, 0.203]
P*	0.438	0.276	0.419	0.172	0.980	0.207	0.450
	(0.261)	(0.265)	(0.315)	(0.315)	(0.240)	(0.311)	(0.306)
	[0.041, 0.886]	[0.037, 0.904]	[0.042, 0.978]	[0.031, 0.978]	[0.237, 0.980]	[0.032, 0.973]	[0.028, 0.960]
Bias(P*)	-0.001	0.000	0.031	5.517	2.191	0.093	0.120
	(0.649)	(0.037)	(0.024)	(5.221)	(3.062)	(0.090)	(0.084)
	[-0.005, 1.741]	[-0.000, 0.101]	[-0.000, 0.067]	[-0.031, 15.368]	[-0.497, 8.139]	[-0.000, 0.259]	[-0.000, 0.239]

Notes: Treatment is defined as participation in SBP. Standard errors in parenthesis and 90% empirical confidence intervals obtained using 500 bootstrap repetitions. OLS = ordinary least squares; BVN = Heckman bivariate normal selection model; HI = Hirano and Imbens (2001) normalized estimator; MB = minimum biased estimator using θ = 0.25, 0.10, 0.05, 0.03, or 0.01.