Estimation of Treatment Effects Without an Exclusion Restriction: with an Application to the Analysis of the School Breakfast Program

Daniel L. Millimet*
Southern Methodist University & IZA

Rusty Tchernis
Georgia State University & NBER

November 3, 2009

Abstract

While the rise in childhood obesity is clear, the policy ramifications are not. School nutrition programs such as the School Breakfast Program (SBP) have come under much scrutiny. However, the lack of experimental evidence, combined with non-random selection into these programs, makes identification of the causal effects of such programs difficult. In the case of the SBP, this difficulty is exacerbated by the apparent lack of exclusion restrictions. Here, we compare via Monte Carlo study several existing estimators that do not rely on exclusion restrictions for identification. In addition, we propose two new estimation strategies. Simulations illustrate the usefulness of our new estimators, as well as provide applied researchers several practical guidelines when analyzing the causal effects of binary treatments. More importantly, we find consistent evidence of a beneficial causal effect of SBP participation on childhood obesity when applying estimators designed to circumvent selection on unobservables.

JEL Classifications: C21, C52, I18, J13
Key words: Treatment Effects, Propensity Score, Unconfoundedness, Selection on Unobservables, Obesity

*The authors are grateful for comments from Jay Bhattacharya, Chris Bollinger, Ozkan Eren, Juan Carlos Escanciano, James Heckman, Keisuke Hirano, Michael Lechner, Arthur Lewbel, Salvador Navarro-Lozano, Denis Nekipelov, Jeff Smith, Aaron Yelowitz, seminar participants at SMU and U Kentucky, and conference participants at the 2009 Econometric Society Winter Meetings and the Fifth IZA Conference on Labor Market Policy Evaluation. Mehtabul Azam provided valuable research assistance. A previous version of this paper was circulated under the title "Minimizing Bias in Selection on Observables Estimators When Unconfoundness Fails." Stata code to implement the estimators used is available from the authors. Corresponding author: Daniel L. Millimet, Department of Economics, Southern Methodist University, Box 0496, Dallas, TX 75275-0496, USA; Email: millimet@mail.smu.edu; Tel. (214) 768-3269; Fax: (214) 768-1821.

1 Introduction

Empirical researchers in economics and other disciplines are often interested in the causal effect of a binary treatment on an outcome of interest. Often randomization is used to ensure comparability (at least in expectation) across the treatment and control groups. However, when a random experiment is not feasible – either due to ethical considerations or cost – researchers must rely on non-experimental or observational data. With such data, non-random selection of subjects into the treatment group becomes a paramount concern.

When subjects self-select into the treatment group on the basis of attributes observable to the researcher, there exist many statistical methods appropriate for the estimation of the causal effect of the treatment. Indeed, the econometric literature on program evaluation has witnessed profound growth in this area over the past few decades.[1] However, when subjects self-select into the treatment group on the basis of attributes unobserved to the researcher, but correlated with the outcome of interest, the situation becomes more difficult for the researcher.

The typical strategy in such circumstances is to rely on an instrumental variable (IV) to identify the causal effect of the treatment. However, a valid instrument is often unavailable. Moreover, even if one is available, it may identify an economically uninteresting parameter in the presence of heterogeneous treatment effects (Imbens and Angrist 1994).

This is the situation we confront in this paper. Here, we are interested in identifying the causal effect of participation in the School Breakfast Program (SBP) on childhood obesity. However, as discussed below, we believe that students in the treatment and control groups differ along important observable and unobservable dimensions. Moreover, we do not believe we have access to a strong, credible instrument. As a result, the usual approach for dealing with non-random selection into SBP – IV using an excluded instrument – does not seem viable.

To proceed, we utilize two existing estimators that do not rely on an exclusion restriction for identification: Heckman's bivariate normal selection estimator that relies on functional form assumptions for identification and a recent IV estimator proposed by Klein and Vella (2009) that exploits heteroskedasticity for identification. In addition, we also propose two new estimators for the analysis of binary treatments where selection into a treatment is non-random, but one lacks an excludable instrument. The first estimator entails minimizing the bias when estimating the effect of a treatment using an estimator that requires conditional independence – independence between treatment assignment and potential outcomes conditional on observables – for unbiasdedness. While minimizing the bias still produces a biased estimate

[1] See Imbens (2004) and Imbens and Wooldridge (2009) for excellent surveys.

of the causal effect of the treatment, our second estimator is a bias-corrected version of the first estimator.

Before applying these methods to the analysis of the causal effect of SBP participation on child health, we first conduct a fairly extensive Monte Carlo study of the estimators considered. This is useful not only in the present context, but also yields several recommendations for practitioners interested in applying these methods to the evaluation of other treatments.

Our Monte Carlo results yield several striking findings. First and foremost, researchers ought to be skeptical of estimates obtained using methods identified from functional form assumptions unless the correct functional form is known. Second, when the appropriate functional form is under-specified in that relevant regressors are omitted from the model, our minimum-biased estimator improves upon the performance of commonly used estimators that require conditional independence, as well as those relying on functional form or heteroskedasticity for identification. Third, when the appropriate functional form is known (or is over-specified in that irrelevant regressors are included in the model), our bias-corrected estimator does quite well, even outperforming estimators relying purely on functional form for identification in some situations. Finally, the Klein and Vella (2009) estimator exploiting heteroskedasticity performs very well when the error in the treatment equation is in fact heteroskedastic, but less well if this is not the case. The Klein and Vella (2009) estimator is also more sensitive to over-specifying the model, whereas our bias-corrected estimator performs better when the model is over-specified.

In terms of the application, our new estimators prove to be a nice complement to existing methods when conditional independence fails, but one lacks an exclusion restriction. Specifically, we find a positive and statistically significant association between SBP and child weight when using estimators that require conditional independence. The relationship remains positive, but becomes statistically insignificant, when we use our minimum-biased estimator. Finally, consistent with prior suggestive evidence in the literature, we find a negative and sometimes statistically significant causal effect of SBP participation on child weight using the bivariate normal selection model and Klein and Vella's (2009) estimator, as well as our bias-corrected estimator. This pattern of point estimates highlights the importance of controlling for non-random selection in studies of the efficacy of the SBP, as well as the important role the SBP plays in the battle against childhood obesity.

The remainder of the paper is organized as follow. Section 2 provides a brief discussion of childhood obesity in the U.S., as well as the relevant literature on school meal programs. Section 3 begins by providing a quick overview of the potential outcomes and corresponding treatment effects framework. Next, it contains our analysis of the bias of estimators that require conditional independence. Finally, we present our new estimation methods, as well as the discuss the existing BVN and KV estimators, in Section 3. Section 4 presents the Monte Carlo study. Section 5 contains our analysis of the SBP. Section

6 concludes.

2 Background

As is quite evident from recent media reports, childhood obesity is deemed to have reached epidemic status (Rosin 2008). Data from the National Health and Nutrition Examination Surveys (NHANES) covering 1976–1980 and 2003–2006 indicate that the prevalence of overweight preschool-aged children, aged 2-5 years, increased from 5.0% to 12.4%.[2] Among school-aged children, the prevalence has risen from 6.5% to 17.0% for those aged 6-11; 5.0% to 17.6% for those aged 12-19 years.[3]

This rise has enormous, long-run implications given that roughly one-third of overweight preschool-aged children and one-half of overweight school-aged children become obese adults (Serdula et al. 1993). In turn, adult obesity is associated with numerous health and socio-economic problems (Bleich et al. 2007). Trasande et al. (2009) estimate that childhood obesity resulted in $237.6 million in hospitalization costs alone in 2005, up (in real terms) from $125.9 million in 2001. Finkelstein et al. (2009) estimate that the total medical costs from obesity – adults and children – exceeded $100 billion in 2006 and could have been as high as $147 billion in 2008.

Given this backdrop, policymakers in the US have acted in a number of different directions, particularly within schools. These reforms culminated in November 2007 as the US Department of Health and Human Services (HHS) launched the Childhood Overweight and Obesity Prevention Initiative. In addition to these recent policy developments, two long-standing federal programs affecting roughly 30 million students on a typical school day have come under scrutiny: the SBP and the National School Lunch Program (NSLP).

The existing literature analyzing these programs – either in combination or in isolation – relies on non-experimental data where the potential exists for non-random selection into treatment on the basis of student-specific unobservable attributes. Three noteworthy approaches have been employed in an attempt to circumvent this issue. First, Bhattacharya et al. (2006) employ a difference-in-difference (DD) strategy and compare the weight and nutritional intake of children during the summer to that during the school year across schools that do and do not offer breakfast. The authors find evidence that children do fare better during the school year along certain dimensions. Second, Schanzenbach (2008) utilizes a regression discontinuity (RD) approach that exploits the sharp income cut-off for eligibility for reduced-price meals to assess the impact of the NSLP. She finds that NSLP participation increases the probability of being obese due to the additional calories provided by school lunches. Finally, Millimet et al. (2009; hereafter

[2] Overweight is defined as an age- and gender-specific body mass index (BMI) greater than the 95th percentile based on growth charts from the Center for Disease Control (CDC).

[3] See http://www.cdc.gov/obesity/childhood/index.html.

MTH) apply the methodology developed in Altonji et al. (2005) to assess the sensitivity of the effects of participation in the SBP and NSLP estimated under exogeneity to deviations from this assumption. The authors find strong, albeit suggestive, evidence that the SBP is beneficial for children, whereas the NSLP is harmful.

Although informative, each of these approaches has limitations. The DD strategy assumes that schools do not decide to participate in the SBP on the basis of the weight trajectories of its students, and that children self-selecting into schools that participate in the SBP do not have different trends in weight and nutrition relative to students attending schools not offering breakfast. In addition, the treatment is school-level breakfast availability, not student-level SBP participation. The RD approach potentially confounds the effects of participation in the SBP and NSLP since the same eligibility criteria is used for both programs. Moreover, the RD approach estimates the local average treatment effect (LATE) when the treatment effect is heterogenous.[4] Finally, the Altonji et al. (2005) approach is merely suggestive, and fails to provide point estimates of the causal effects of the programs except under functional form assumptions.

In this study, we assess the causal role of the SBP in the child obesity epidemic, as well as propose new techniques applicable more generally to the evaluation of binary treatments. We focus on the SBP for two reasons. First, MTH and Schanzenbach (2008) both find little evidence that NSLP participation is non-random conditional on observables. Second, MTH finds evidence of *positive selection* into the SBP: children with steeper weight trajectories over time are more likely to participate conditional on observables. Moreover, addressing non-random selection into the SBP is vital not only for obtaining consistent estimates of the causal effect of SBP participation, but also for estimating the causal effect of NSLP participation.[5] MTH conclude that *if* there is positive selection into SBP participation on child weight, *then* the SBP lowers child weight *and* the NSLP raises child weight once this non-random selection is addressed. Because the underlying selection into SBP is crucial for consistently estimating the causal effects of both programs, and the analysis in MTH is only suggestive of a beneficial impact of SBP participation once selection is addressed, obtaining point estimates of the causal effect of SBP participation utilizing different techniques is of critical importance to policymakers.

[4] See Lee and Lemieux (2009) for a detailed discussion.

[5] This is true despite MTH concluding that there is no selection into NSLP on the basis of student-level unobservables. The reason is that if SBP is erroneously treated as exogenous, then the bias due to correlation with the error term spills over and also biases the coefficient on NSLP participation since SBP and NSLP participation are highly correlated.

3 The Evaluation Problem

3.1 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,

$$\tau_i = Y_i(1) - Y_i(0).$$

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

$$\begin{aligned} \tau_{ATE} &= \mathrm{E}[\tau_i] = \mathrm{E}[Y_i(1) - Y_i(0)] && (1) \\ \tau_{ATT} &= \mathrm{E}[\tau_i | T = 1] = \mathrm{E}[Y_i(1) - Y_i(0) | T = 1] && (2) \\ \tau_{ATU} &= \mathrm{E}[\tau_i | T = 0] = \mathrm{E}[Y_i(1) - Y_i(0) | T = 0]. && (3) \end{aligned}$$

In general, the parameters in (1) – (3) 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:

$$\begin{aligned} \tau_{ATE}[X] &= \mathrm{E}[\tau_i | X] = \mathrm{E}[Y_i(1) - Y_i(0) | X] && (4) \\ \tau_{ATT}[X] &= \mathrm{E}[\tau_i | X, T = 1] = \mathrm{E}[Y_i(1) - Y_i(0) | X, T = 1] && (5) \\ \tau_{ATU}[X] &= \mathrm{E}[\tau_i | X, T = 0] = \mathrm{E}[Y_i(1) - Y_i(0) | X, T = 0]. && (6) \end{aligned}$$

The parameters in (1) – (3) are obtained by taking the expectation of the corresponding parameter in (4) – (6) 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

$$Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$$

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.

The methods utilized by researchers to circumvent this missing data problem are classified into two groups: selection on observables estimators and selection on unobservable estimators. The distinction lies in whether a method consistently estimates the causal effect of the treatment in the presence of unobservable attributes of subjects that are correlated with both treatment assignment and the outcome of interest conditional on the set of observable variables. Assuming a lack of such unobservables is referred to as *conditional independence* or *unconfoundedness* assumption (Rubin 1974; Heckman and Robb 1985). Formally, under the conditional independence assumption (CIA), 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.

If CIA fails to hold, then consistent estimation of the causal effect requires a selection on unobservables estimation technique. The difficulty for applied researchers confronting such a scenario is that obtaining a consistent *point estimate* of a measure of the treatment effect typically requires an exclusion restriction (i.e., an observable variable that impacts treatment assignment, but not the outcome of interest conditional on treatment assignment).

Unfortunately, in many situations, valid exclusion restrictions, as usually conceived, are not available. At that point, researchers have limited options: (i) abandon point estimation, instead focusing on bounding a measure of the treatment effect, (ii) rely on strong functional form assumptions for identification, or (iii) identify the causal effect from higher moments of the observed variables or heteroskedasticity. Here, we propose two new options in addition to applying previous estimators utilizing functional form assumptions and heteroskedasticity for identification. First, we present a *minimum-biased* estimator. This approach entails the utilization of a selection on observables estimator, but restricts the estimation sample in such a way so as to minimize the bias arising from the failure of CIA. Second, we extend our minimum-biased estimation approach by suggesting a *bias-corrected* version of this estimator.

To proceed, we first derive the bias of estimators that require CIA when, in fact, this assumption fails under certain assumptions (namely, joint normality). We then derive our minimum-biased and bias-corrected estimators under these same assumptions. Finally, we extend our estimators to the case where joint normality does not hold.

3.2 Bias When CIA 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 CIA.[6] Regardless of which such technique is employed, each will be biased if CIA fails to hold.

Black and Smith (2004) and Heckman and Navarro-Lozano (2004) consider the bias when estimating the ATT under CIA and the assumption is incorrect. The bias of the ATT at some value of the propensity score, $P(X)$, is given by

$$\begin{aligned} B_{ATT}[P(X)] &= \widehat{\tau}_{ATT}[P(X)] - \tau_{ATT}[P(X)] \\ &= \{\mathrm{E}[Y(1)|T=1, P(X)] - \mathrm{E}[Y(0)|T=0, P(X)]\} \\ &\quad - \{\mathrm{E}[Y(1)|T=1, P(X)] - \mathrm{E}[Y(0)|T=1, P(X)]\} \\ &= \mathrm{E}[Y(0)|T=1, P(X)] - \mathrm{E}[Y(0)|T=0, P(X)] \end{aligned} \tag{7}$$

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

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

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

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

[6] D'Agostino (1998) and Imbens (2004) offer excellent surveys.

(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, (7) simplifies to

$$\begin{aligned} B_{ATT}[P(X)] &= \mathrm{E}[\varepsilon_0 | T = 1, P(X)] - \mathrm{E}[\varepsilon_0 | T = 0, P(X)] \\ &= -\rho_{0u}\sigma_0 \frac{\phi(h(X))}{\Phi(h(X))[1 - \Phi(h(X))]} \end{aligned} \tag{8}$$

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 (2). Indeed, the procedure suggested in Black and Smith (2004) accomplishes the following. It searches over the parameters defined in (5) to find the value of $P(X)$ for which the $\tau_{ATT}[P(X)]$ can be estimated with the least bias. This 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

$$\begin{aligned} B_{ATU}[P(X)] &= \mathrm{E}[\varepsilon_1 | T = 1, P(X)] - \mathrm{E}[\varepsilon_1 | T = 0, P(X)] \\ &= -\rho_{1u}\sigma_1 \frac{\phi(h(X))}{\Phi(h(X))[1 - \Phi(h(X))]} \end{aligned} \tag{9}$$

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 $P(X)$.[7]

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

$$\begin{aligned} B_{ATE}[P(X)] &= -\rho_{0u}\sigma_0 \left\{ \frac{\phi(h(X))}{\Phi(h(X))[1 - \Phi(h(X))]} \right\} + [1 - P(X)] \left\{ -\rho_{\delta u}\sigma_\delta \frac{\phi(h(X))}{\Phi(h(X))[1 - \Phi(h(X))]} \right\} \\ &= -\left\{ \rho_{0u}\sigma_0 + [1 - P(X)]\rho_{\delta u}\sigma_\delta \right\} \left\{ \frac{\phi(h(X))}{\Phi(h(X))[1 - \Phi(h(X))]} \right\} \end{aligned} \tag{10}$$

which is equivalent to the expression derived in Heckman and Navarro-Lozano (2004).

Equation (10) leads to three salient points. First, under A1 – A3, $|B_{ATU}[P(X)]| \geqslant |B_{ATE}[P(X)]| \geqslant |B_{ATT}[P(X)]|$. Second, the value of $P(X)$ that minimizes the bias of the ATE, referred to as the *bias minimizing propensity score* (BMPS) and denoted $P^*(X)$, is not fixed; rather, it depends on the values of $\rho_{0u}\sigma_0$ and $\rho_{\delta u}\sigma_\delta$. In particular, 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). Third, there are two special cases for which the bias disappears:

(i) No Selection on Unobservables Impacting Outcomes in the Untreated State: If $\rho_{0u}\sigma_0 = 0$, then $\lim_{P(X)\to 1} B_{ATE}[P(X)] = 0$

(ii) Equal but Opposite Selection on Unobservables: If $\rho_{0u}\sigma_0 = -\rho_{\delta u}\sigma_\delta$, then $\lim_{P(X)\to 0} B_{ATE}[P(X)] = 0$

Beyond these cases, $P^*(X)$ displays the following properties:

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

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

(a) above 0.5, but monotonically increasing in $\rho_{\delta u}\sigma_\delta > 0$

(b) above or below 0.5 for $\rho_{\delta u}\sigma_\delta < 0$, but strictly above 0.5 for $\rho_{0u}\sigma_0 < -0.5\rho_{\delta u}\sigma_\delta$ (i.e., strong, positive selection on unobserved gains)

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

(a) above 0.5, but monotonically decreasing in $\rho_{\delta u}\sigma_\delta < 0$

(b) above or below 0.5 for $\rho_{\delta u}\sigma_\delta > 0$, but strictly above 0.5 for $\rho_{0u}\sigma_0 > -0.5\rho_{\delta u}\sigma_\delta$ (i.e., strong, negative selection on unobserved gain

Interestingly, when the two types of selection operate in same direction, $P^*(X) > 0.5$ implying that the treatment effect for individuals with a higher probability of receiving the treatment can be estimated with less bias. However, when these two types of selection go in opposite direction, $P^*(X) > 0.5$ when selection on unobserved gains is relatively large in absolute value. Thus, in this case, when there is relatively strong selection on unobserved gains, the treatment effect for individuals with a higher probability of receiving the treatment can be estimated with less bias. Figure 1 provides a complete characterization of $P^*(X)$.

4 Estimation

4.1 The Minimum-Biased Approach

In light of the preceding discussion, we propose a new estimation approach when unconfoundedness is not likely to hold, but one lacks a valid exclusion restriction. We couch our technique within the normalized inverse probability weighted estimator of Hirano and Imbens (2001).[8] First, consider estimation of the ATE. Horvitz and Thompson (1952) show that the ATE may be expressed as

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

with the sample analogue given by

$$\hat{\tau}_{HT,ATE} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{Y_i T_i}{\hat{P}(X_i)} - \frac{Y_i(1-T_i)}{1-\hat{P}(X_i)}\right]. \tag{11}$$

[8] Other propensity score estimators, such as matching, could be used as well. Millimet and Tchernis (2009) find little practical difference between estimator of Hirano and Imbens (2001) and kernel matching.

The estimator in (11) 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* or HI estimator, which is given by

$$\hat{\tau}_{HI,ATE} = \left[\sum_{i=1}^{N} \frac{Y_i T_i}{\hat{P}(X_i)} \Big/ \sum_{i=1}^{N} \frac{T_i}{\hat{P}(X_i)}\right] - \left[\sum_{i=1}^{N} \frac{Y_i(1-T_i)}{1-\hat{P}(X_i)} \Big/ \sum_{i=1}^{N} \frac{(1-T_i)}{1-\hat{P}(X_i)}\right]. \tag{12}$$

Millimet and Tchernis (2009) provide evidence of the superiority of the normalized estimator in practical settings.

Under CIA, the HI estimator in (12) provides an unbiased estimate of τ_{ATE}. When this assumption fails, the bias is given (10). Rather than abandon this estimator, however, we propose to alter the estimator in such a manner so as to minimize the bias. Specifically, we suggest estimating (12) using only observations with a propensity score in a neighborhood around the BMPS, P^*.[9] Formally, we propose the following *minimum-biased* (MB) estimator of the ATE:

$$\hat{\tau}_{MB,ATE}[P^*] = \left[\sum_{i\in\Omega} \frac{Y_i T_i}{\hat{P}(X_i)} \Big/ \sum_{i\in\Omega} \frac{T_i}{\hat{P}(X_i)}\right] - \left[\sum_{i\in\Omega} \frac{Y_i(1-T_i)}{1-\hat{P}(X_i)} \Big/ \sum_{i\in\Omega} \frac{(1-T_i)}{1-\hat{P}(X_i)}\right] \tag{13}$$

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 $\theta = 0.05$ and 0.25. For example, if $\theta = 0.05$, $\alpha_{0.05}$ is the smallest value 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

[9] For simplicity, we suppress the notation denoting the fact that the BMPS depends on X.

Heckman's 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

$$y_i = X_i\beta_0 + X_iT_i(\beta_1 - \beta_0) + \beta_{\lambda 0}(1 - T_i)\left[\frac{\phi(X_i\gamma)}{1 - \Phi(X_i\gamma)}\right] + \beta_{\lambda 1}T_i\left[\frac{-\phi(X_i\gamma)}{\Phi(X_i\gamma)}\right] + \eta_i \tag{14}$$

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

$$\begin{aligned} \beta_{\lambda 0} &= \rho_{0u}\sigma_0 \\ \beta_{\lambda 1} &= \rho_{0u}\sigma_0 + \rho_{\delta u}\sigma_\delta. \end{aligned} \tag{15}$$

Thus, OLS estimation of (14) 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 (10) to obtain an estimate of P^*.[10]

Our proposed estimator immediately raises a question: If one is willing to maintain the assumptions underlying the BVN selection model in (14), why not just use the OLS estimates of (14) to estimate the ATE? Well, perhaps one should. Upon estimating the BVN selection model, the ATE is given by

$$\hat{\tau}_{BVN,ATE} = \overline{X}\left(\widehat{\beta}_1 - \widehat{\beta}_0\right). \tag{16}$$

However, if the assumptions of the BVN model do not hold, or if the BVN model is poorly identified, then perhaps the estimator in (12) or (13) will perform better in practice. Moreover, when one is interested in the ATT rather than the ATE, our minimum biased approach can be utilized without estimation of (14) since P^* is known to be one-half. In this case, our estimator – couched in the HI estimator of the ATT – is given by

$$\hat{\tau}_{MB,ATT}[P^*] = \sum_{i\in\Omega} Y_iT_i - \left[\sum_{i\in\Omega}\frac{Y_i(1-T_i)\hat{P}(X_i)}{1-\hat{P}(X_i)} \Bigg/ \sum_{i\in\Omega}\frac{(1-T_i)\hat{P}(X_i)}{1-\hat{P}(X_i)}\right]. \tag{17}$$

For comparison, under the functional form assumptions for $g_0(X)$, $g_1(X)$, and $h(X)$ discussed above,

[10] 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.

OLS estimation of (14) also produces a consistent estimate of the ATT. This estimator is given by

$$\hat{\tau}_{BVN,ATT} = \overline{X}_1 \left(\widehat{\beta}_1 - \widehat{\beta}_0\right) + \widehat{\beta}_{\lambda 1} \overline{\left[\frac{-\phi(X_i\widehat{\gamma})}{\Phi(X_i\widehat{\gamma})}\right]_1} \tag{18}$$

where $\overline{X}_1$ and $\overline{\left[\frac{-\phi(X_i\widehat{\gamma})}{\Phi(X_i\widehat{\gamma})}\right]_1}$ are the sample means of X and the selection correction term, respectively, in treatment group.

4.2 The Bias-Corrected Approach

While (13) and (17) provide a minimum-biased estimator of the ATE and ATT, respectively, estimation of the error correlation structure using (14) immediately gives rise to the possibility of a bias-corrected version of each estimator. Specifically, after estimating (14), estimates of the *bias* of the minimum-biased estimator of the ATE and ATT are given by

$$\widehat{B_{ATE}[P^*]} = -\left[\widehat{\rho_{0u}\sigma_0} + (1-P^*)\widehat{\rho_{\delta u}\sigma_\delta}\right]\left[\frac{\phi(\Phi^{-1}(P^*))}{P^*(1-P^*)}\right]$$

and

$$\widehat{B_{ATT}[0.5]} \approx -1.6 * \widehat{\rho_{0u}\sigma_0}.$$

Bias-corrected estimators, denoted MB-BC, are then given by

$$\hat{\tau}_{MB-BC,ATE}[P^*] = \widehat{\tau}_{MB,ATE}[P^*] - \widehat{B_{ATE}[P^*]} \tag{19}$$

$$\hat{\tau}_{MB-BC,ATT}[P^*] = \widehat{\tau}_{MB,ATT}[P^*] - \widehat{B_{ATT}[P^*]}. \tag{20}$$

4.3 Deviations from Normality

The analysis to this point has relied on the assumption of joint normality; an assumption that may be unrealistic in many applications. As a result, before turning to our Monte Carlo study, we consider deviations from the normality assumption. To proceed, we follow Lee (1984) who utilizes the fact that under certain assumptions the (unknown) joint density of ε_j and u, $f_j(\varepsilon_j, u)$, $j = 0, 1$, may be written as a bivariate Edgeworth series of distributions. Formally,

$$f_j(\varepsilon_j, u) = \phi_2(\varepsilon_j, u) + \sum_{r+s\geq 3} (-1)^{r+s} A_{rs} \frac{1}{r!s!} \frac{\partial^{r+s}\phi_2(\varepsilon_j, u)}{\partial u^r \partial \varepsilon_j^s} \tag{21}$$

where ϕ_2 is the bivariate standard normal density and A_{rs} are functions of the cumulants (or semi-invariants) of ε_j and u.

As in Lee (1984), we consider the case where $r+s \in \{3,4\}$, referred to as a Type AA surface in Mardia (1970), and $u \sim N(0,1)$. In light of this, the bias of the ATT is now given by

$$\begin{aligned} B_{ATT}[P(X)] &= \mathrm{E}[\varepsilon_0|T=1,P(X)] - \mathrm{E}[\varepsilon_0|T=0,P(X)] \\ &= -\left\{\rho_{0u}\sigma_0 + \kappa_{12}\sigma_0\frac{h(X)}{2} + \kappa_{13}\sigma_0\frac{\left[h(X)^2-1\right]}{6}\right\}\left\{\frac{\phi(h(X))}{\Phi(h(X))[1-\Phi(h(X))]}\right\} \end{aligned} \tag{22}$$

where κ_{ij} are semi-invariants (or cumulants). For the ATU, we have by symmetry

$$\begin{aligned} B_{ATU}[P(X)] &= \mathrm{E}[\varepsilon_1|T=1,P(X)] - \mathrm{E}[\varepsilon_1|T=0,P(X)] \\ &= -\left\{\rho_{1u}\sigma_1 + \kappa'_{12}\sigma_1\frac{h(X)}{2} + \kappa'_{13}\sigma_1\frac{\left[h(X)^2-1\right]}{6}\right\}\left\{\frac{\phi(h(X))}{\Phi(h(X))[1-\Phi(h(X))]}\right\} \end{aligned} \tag{23}$$

Alternatively,

$$\begin{aligned} B_{ATU}[P(X)] &= \mathrm{E}[\delta+\varepsilon_0|T=1,P(X)] - \mathrm{E}[\delta+\varepsilon_0|T=0,P(X)] \\ &= B^{ATT}[P(X)] + \mathrm{E}[\delta|T=1,P(X)] - \mathrm{E}[\delta|T=0,P(X)] \\ &= -\left\{\begin{array}{c} \rho_{0u}\sigma_0 + \kappa_{12}\sigma_0\frac{h(X)}{2} + \kappa_{13}\sigma_0\frac{\left[h(X)^2-1\right]}{6} \\ +\rho_{\delta u}\sigma_\delta + \kappa''_{12}\sigma_\delta\frac{h(X)}{2} + \kappa''_{13}\sigma_\delta\frac{\left[h(X)^2-1\right]}{6} \end{array}\right\}\left\{\frac{\phi(h(X))}{\Phi(h(X))[1-\Phi(h(X))]}\right\} \end{aligned}$$

For the ATE, the bias is

$$B_{ATE}[P(X)] = -\left\{\begin{array}{c} \rho_{0u}\sigma_0 + \kappa_{12}\sigma_0\frac{h(X)}{2} + \kappa_{13}\sigma_0\frac{\left[h(X)^2-1\right]}{6} \\ +[1-P(X)]\left\{\rho_{\delta u}\sigma_\delta + \kappa''_{12}\sigma_\delta\frac{h(X)}{2} + \kappa''_{13}\sigma_\delta\frac{\left[h(X)^2-1\right]}{6}\right\} \end{array}\right\}\left\{\frac{\phi(h(X))}{\Phi(h(X))[1-\Phi(h(X))]}\right\} \tag{24}$$

Minimizing the bias of the three parameters requires knowledge or estimates of several additional parameters. However, these are estimable from an altered version of the BVN selection model:

$$\begin{aligned} y_i &= X_i\beta_0 + X_iT_i(\beta_1-\beta_0) \\ &+\beta_{\lambda01}(1-T_i)\left[\frac{\phi(X_i\gamma)}{1-\Phi(X_i\gamma)}\right] + \beta_{\lambda02}(1-T_i)\left[\frac{X_i\gamma}{2}\frac{\phi(X_i\gamma)}{1-\Phi(X_i\gamma)}\right] + \beta_{\lambda03}(1-T_i)\left[\frac{(X_i\gamma)^2-1}{6}\frac{\phi(X_i\gamma)}{1-\Phi(X_i\gamma)}\right] \\ &+\beta_{\lambda11}T_i\left[\frac{-\phi(X_i\gamma)}{\Phi(X_i\gamma)}\right] + \beta_{\lambda12}T_i\left[\frac{-X_i\gamma}{2}\frac{\phi(X_i\gamma)}{\Phi(X_i\gamma)}\right] + \beta_{\lambda13}T_i\left[\frac{1-(X_i\gamma)^2}{6}\frac{\phi(X_i\gamma)}{\Phi(X_i\gamma)}\right] + \eta_i \end{aligned} \tag{25}$$

where

$$\begin{aligned}
\beta_{\lambda 01} &= \rho_{0u}\sigma_0 \\
\beta_{\lambda 02} &= \kappa_{12}\sigma_0 \\
\beta_{\lambda 03} &= \kappa_{13}\sigma_0 \\
\beta_{\lambda 11} &= \rho_{0u}\sigma_0 + \rho_{\delta u}\sigma_\delta \\
\beta_{\lambda 12} &= \kappa_{12}\sigma_0 + \kappa''_{12}\sigma_\delta \\
\beta_{\lambda 13} &= \kappa_{13}\sigma_0 + \kappa''_{13}\sigma_\delta.
\end{aligned}$$

Upon estimation of (25), the BMPS is found by minimizing the bias in (24). The remainder of the estimation algorithm is unchanged.

4.4 Klein & Vella (2009) Estimator

For comparison to the previous estimators, we also employ the estimator of Klein and Vella (2009; hereafter KV). Our parametric implementation of this estimator relies on a similar functional form assumption to the BVN estimator in the absence of heteroskedasticity, but effectively induces valid exclusion restrictions in the presence of heteroskedasticity. To proceed, suppose that latent treatment assignment is now given by

$$T^* = X\gamma - u^*$$

where $u^* = S(X)u$ and u is drawn from a standard normal density. In this case, the probability of receiving the treatment conditional on X is given by

$$\Pr(T = 1|X) = \Phi\left(\frac{X}{S(X)}\gamma\right). \tag{26}$$

Assuming $S(X) = \exp(X\delta)$, the parameters of (26) are estimable by maximum likelihood (ML), with the log-likelihood function given by

$$\ln \mathcal{L} = \sum\nolimits_i \left[\ln \Phi\left(\frac{X\gamma}{\exp(X\delta)}\right)\right]^{T_i} \left\{\ln\left[1 - \Phi\left(\frac{X\gamma}{\exp(X\delta)}\right)\right]\right\}^{1-T_i} \tag{27}$$

where the element of δ corresponding to the intercept is normalized to zero for identification.

The ML estimates are then used to obtain the predicted probability of treatment, $\widehat{P(X)}$, which may be used as an instrument for T in equation (14) excluding the selection correction terms.[11] Note, even if

[11] Interactions between $\widehat{P(X)}$ and X may also serve as instruments given inclusion of interactions between T and X in (14).

$S(X) = 1$, $\widehat{P(X)}$ remains a valid instrument since it is non-linear in X. However, since the non-linearity arises mostly in the tails, identification typically relies on a small fraction of the sample. On the other hand, if $S(X) \neq 1$, then the KV approach effectively induces a valid exclusion restriction as $Z \equiv X/S(X)$ is frequently linearly independent of X (Klein and Vella 2009).

5 Monte Carlo Study

5.1 Setup

To assess the performance of the various estimators, we use two primary experimental designs. In the first design, the data-generating process (DGP) imposes the constant treatment effect setup (i.e., $\tau_i = \tau$ for all i). In the second design, the DGP gives rise to heterogeneous treatment effects, where the heterogeneity is due to unobserved, individual-specific gains to treatment (i.e., τ_i varies across i, but this variation arises due to differences in $\varepsilon_{1i} - \varepsilon_{0i}$). In this study, in part motivated by our application, we do not consider the case where the treatment effect is heterogeneous on the basis of observables; the vectors β_0 and β_1 are identical except for the intercept.

We assess finite sample performance by simulating 250 data sets, each with 5,000 observations, containing

$$x_1, x_2 \overset{\text{iid}}{\sim} U(-1, 1)$$

and

$$\begin{aligned} h(X) &= 0.5(x_1 - x_2) + 0.5(x_1^2 - x_2^2) + 2x_1x_2 \\ T^* &= 0.5 + h(X) - u \\ T &= \begin{cases} 1 & \text{if } T^* > 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

where u is defined below.[12]

In the constant treatment effect setup, there is a common effect $\varepsilon = \varepsilon_0 = \varepsilon_1$ and potential outcomes are given by

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

[12] We also simulated 50 data sets with 250,000 observations each to assess large sample performance. Results are available in Appendix B.

which implies that $\tau_i = 1$ for all i and $\tau_{ATE} = \tau_{ATT} = \tau_{ATU} = 1$. In the second design, where treatment effects are heterogeneous on the basis of unobserved gains, potential outcomes are given by

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

which implies that $\tau_i = 1 + \varepsilon_{1i} - \varepsilon_{0i} = 1 + \delta_i$ and $\tau_{ATE} = 1 + \mathrm{E}[\delta_i]$, $\tau_{ATT} = 1 + \mathrm{E}[\delta_i | T_i = 1]$, and $\tau_{ATU} = 1 + \mathrm{E}[\delta_i | T_i = 0]$.

Within each of the two experimental designs, we consider four general error structures. In the first structure, the errors are multivariate normal

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

where

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

and $\rho_{01} = 1$ in the common effect setup, implying $\rho_{\delta u} = 0$ and $\rho_{0u} = \rho_{1u}$. In the heterogeneous effect setup, $\rho_{01} = 0.5$ implying $\rho_{\delta u} = \rho_{1u} - \rho_{0u}$. In the second error structure, the errors are drawn from an asymmetric and non-normal multivariate distribution with the desired correlation matrix using the method in Headrick and Sawilowsky (1999). Specifically, ε_0 and ε_1 are mean zero with unit variance and skewness and kurtosis close to a χ^2_1 distribution (skewness is roughly $\sqrt{8}$ and kurtosis is around 15); u continues to have a standard normal distribution. The third and fourth error structures are identical to the preceding cases except we introduce heteroskedasticity in the treatment assignment equation. Specifically, we define $u^* = (1 + 0.45(x_1 + x_2))u$ and $T^* = 0.5 + h(X) - u^*$.

In the constant treatment effect experimental design, we consider three cases: ρ_{0u} equal to zero, -0.25, and -0.50, where CIA holds in the first case and larger values of ρ_{0u} (in absolute value) correspond to greater selection on unobservables.[13] In the heterogeneous treatment effect experimental design, we consider three cases: $\rho_{0u} = \rho_{\delta u} = 0$, $\rho_{0u} = -0.20$ and $\rho_{\delta u} = -0.10$, and $\rho_{0u} = -0.40$ and $\rho_{\delta u} = -0.10$, where CIA again holds in the first case and larger values of ρ_{0u} (in absolute value) correspond to greater selection on unobservables.

Finally, for each DGP considered, we use four specifications when implementing the various estimators.

[13] Given the DGP for T^*, $\rho_{0u} < 0$ indicates positive selection (i.e., unobservables associated with better outcomes are also associated with a higher probability of receiving the treatment).

The specifications differ in terms of the functional form of the treatment assignment model, as well as the outcome equations in (14) and (25). Specification (1) includes only linear terms for x_1 and x_2 (hence, x_1^2, x_2^2, and x_1x_2 are omitted) in the probit model for treatment assignment and in (14) and (25). Specification (2) includes x_1, x_2, x_1^2, x_2^2, and x_1x_2 in the set of covariates used in the probit model for treatment assignment and in (14) and (25). Specification (3) is identical to Specification (2) but adds cubic terms for x_1 and x_2 as well as the complete set of interaction terms between the linear and quadratic terms. Specification (4) is identical to Specification (3) but incorporates an additional, completely irrelevant regressor, $x_3 \sim U(-1, 1)$, and includes x_3, x_3^2, x_3^3, and the full set of interaction terms in the set of covariates. In Specifications (1) – (3), we model $S(X) = S(x_1, x_2)$; in specification (4), we model $S(X) = S(x_1, x_2, x_3)$.

Two final comments are warranted. First, we assume in all cases that the researcher knows that the treatment effect does not vary on the basis of observables. As such, we exclude interactions between T and the set of covariates in X in (14) and (25). Second, Specification (1) is *under-specified*, Specification (2) is the correct specification, and Specifications (3) and (4) are *over-specified*. Examination of these different cases is motivated by the fact that the applied researcher likely does not know the correct functional form even if the researcher knows the set of covariates in the model. In addition, Millimet and Tchernis (2009) find that over-specifying the propensity score model when unconfoundedness holds is warranted. We explore whether this conclusion extends to the current situation.

5.2 Results

5.2.1 Correct Specification

Results for the ATE and ATT from the first experimental design – constant treatment effect – using the correct specification, Specification (2), are presented in Table 1. Figures represent Root Mean Squared Errors (RMSE). Panel I displays the results when the errors are normal; Panel II displays the results when the errors are asymmetric and non-normal. The left (right) set of columns corresponds to the case where the error in the treatment assignment is homoskedastic (heteroskedastic). Within each column, the shaded number represents the smallest RMSE.

Columns (1) and (4) correspond to the simplest case: the treatment effect is constant and conditional independence holds. Three findings emerge. First, consonant with our expectations, the HI estimator performs best for the ATE with or without homoskedastic errors (Column (1)). However, our MB estimator with a large radius ($\theta = 0.25$) fares only marginally worse than HI. Second, the HI estimator and MB estimator with a large radius ($\theta = 0.25$) achieve almost identical performance for the ATT when errors are homoskedastic; our MB estimator with a large radius ($\theta = 0.25$) performs marginally worse than HI with

heteroskedastic errors (Column (4)). Third, even when normality holds, KV, BVN, and MB-BC have a higher RMSE than HI and MB for both the ATE and ATT. This is not surprising given the efficiency loss from not utilizing conditional independence when this holds in the DGP. Finally, BVN, MB, and MB-BC outperform their counterparts allowing for non-normality even when the errors are non-normal. Thus, the flexibility offered by these estimators is more than offset by the imprecision resulting from the increased parameterization of the model.

Columns (2), (3), (5), and (6) introduce selection on unobservables into the DGP, with (3) and (6) containing the largest amount of selection on unobservables. For the ATE (Columns (2) and (3)), three salient conclusions emerge. First, with relatively modest selection on unobservables (Column (2)), KV, BVN, and MB-BC with a large radius ($\theta = 0.25$) perform comparably, with KV doing marginally worse than the other two estimators when errors are homoskedastic. Second, with relatively strong selection on unobservables (Column (3)), KV performs slightly better than BVN and MB-BC with a large radius ($\theta = 0.25$) except in the case of normal and homoskedastic errors where BVN and MB-BC have lower RMSEs. Finally, as in the case of selection on observables, BVN and MB-BC outperform their counterparts allowing for non-normality even when the errors are non-normal.

When there is selection on unobservables and one is interested in the ATT (Columns (5) and (6)), the conclusions are only slightly altered. First, with relatively modest selection on unobservables (Column (5)), KV, BVN, and MB-BC with a large radius ($\theta = 0.25$) continue to have similar RMSEs with BVN doing marginally better than the other two estimators except in the case of non-normal and heteroskedastic errors. Second, with relatively strong selection on unobservables (Column (6)), KV, BVN, and MB-BC with a large radius ($\theta = 0.25$) perform comparably when errors are normal; KV performs significantly better than the other two estimators when errors are non-normal. This result holds regardless of whether the errors are homoskedastic or not. Finally, as in the prior cases, BVN and MB-BC outperform their counterparts allowing for non-normality even when normality does not hold.

Next, we turn to the results for the ATE and ATT from the second experimental design. Recall, with this DGP, the treatment effect varies across observations due to individual-specific, unobserved gains to treatment. Results obtained using the correct specification, Specification (2), are presented in Table 2.

Columns (1) and (4) again correspond to the case of selection on observables. The qualitative findings are unchanged from the corresponding columns in Table 1. Specifically, HI continues to perform best for both ATE and ATT, but MB with a large radius ($\theta = 0.25$) has only a marginally higher RMSE. Similarly, the majority of the findings from Table 1 continue to hold when focusing on the ATE and ATT under selection on unobservables (Columns (2), (3), (5), and (6)). The only noteworthy change is that now BVN outperforms KV in the case of non-normal and homoskedastic errors with relatively strong selection on

unobservables for the ATE (Column (3)).[14]

Taking a step back, three conclusions emerge that should be of interest to applied researchers. First, under selection on observables, HI is preferred, but MB with a large radius is very comparable even in small samples. Moreover, MB is more robust to selection on unobservables than HI when the focus is on the ATE. Thus, MB provides researchers with an alternative estimator that can serve as a useful robustness check to the usual HI estimator when the selection on observables assumption is questionable, but the parameter of interest is the ATE. Second, if the researcher is confident there is selection on unobservables, then KV, BVN, and MB-BC with a large radius are the preferred estimators. In particular, BVN is preferred when errors are homoskedastic, unless the focus is on the ATT and there is relatively strong selection on unobservables in which case MB-BC is preferred. KV is preferred when errors are heteroskedastic. As stated earlier, since the MB and MB-BC estimators rely on the BVN estimator, it might seem odd not to simply use the BVN estimator. However, as the simulations indicate, there are multiple scenarios in which the MB or MB-BC estimator outperform the BVN (and KV) estimators. Finally, there is little gain in general from the estimators based on deviations from normality. This arises due to the noise introduced through the estimation of additional parameters in (25).

5.2.2 Under- and Over-specifying the Model

The preceding results assume that the proper specification for the observables is known. Since this is rarely true in practice, we now assess the performance of the estimators considered herein when this assumption is relaxed. The results are relegated to Appendix A.

Tables A1–A3 correspond to the first experimental design where the treatment effect is constant across all observations. Table A1 corresponds to the case of selection on observables, A2 to the case of moderate selection on unobservables, and A3 to the case of strong selection on unobservables. Within each table, Columns (1) through (4) correspond to Specifications (1) – (4) defined previously; thus, Column (1) is under-specified, Column (2) is correctly specified and simply repeats the results from Table 1, and Columns (3) and (4) are over-specified. Tables A4–A6 are similarly organized, but correspond to the second experimental design where the treatment is heterogeneous due to individual-specific, unobserved gains to treatment.

To begin, consider the case of selection on observables when the focus is on the ATE (Table A1 and A4). Four interesting conclusions can be drawn. First, HI continues have the smallest RMSE when the

[14] There are also two instances where MB (Table 2, Panel II, Column (3) with heteroskedastic errors) and MB-EE (Table 2, Panel II, Column (5) with homoskedastic errors) perform best. We do not pay much attention to these cases as these estimators do not significantly outperform the other estimators in these specifications, while they perform particularly poorly in other specifications. Thus, a researcher would be hard-pressed to justify use of these estimators when the exact DGP is unknown.

model is over-fitted. In addition, as demonstrated in Millimet and Tchernis (2009), there is no penalty to over-fitting, and in many cases the RMSE actually falls. This is also consistent with Hirano et al. (2003) who find that using an estimated propensity score is preferable even when the true propensity score is known, as well as recent work in Abadie and Imbens (2009) who find that the asymptotic variance of propensity score matching estimators is smaller when the propensity score is estimated (as opposed to known).

Second, as found in Tables 1 and 2 using the correct specification, MB with a large radius ($\theta = 0.25$) is very comparable in performance to HI and, as with HI, over-fitting improves its performance. Third, if the model is under-specified, MB-EE with a small radius ($\theta = 0.05$) does best when the errors are homoskedastic, whereas MB with a small radius ($\theta = 0.05$) does best when the errors are heteroskedastic. Thus, when the researcher believes that conditional independence holds, but is unsure of the proper functional form, MB with a small radius provides a viable sensitivity check as it performs significantly better than HI and MB-EE when errors are heteroskedastic, and only marginally worse than MB-EE (but still much better than HI) when errors are homoskedastic.

Finally, when the model is under-specified (Column (1)), which induces selection on unobservables, KV, BVN, BVN-EE, MB-BC, and MB-BC-EE perform very poorly despite their design to circumvent selection on unobservables. In fact, the RMSE for KV is roughly eight times larger than MB and MB-EE, while the RMSE for BVN and MB-BC (BVN-EE and MB-BC-EE) is at least 30 (550) times larger. Thus, when applying these estimators, researchers must be very wary of under-specifying the model.

When selection is still only on observables but one focuses on the ATT, the conclusions drawn change somewhat. Specifically, HI continues to do best when the model is over-fitted, with MB with a large radius ($\theta = 0.25$) performing only marginally worse, and over-fitting benefiting the performance of both estimators. In addition, KV, BVN, BVN-EE, MB-BC, and MB-BC-EE continue to perform very poorly when the model is under-specified. However, if the model is under-specified, now MB with a small radius ($\theta = 0.05$) does best when the errors are homoskedastic, whereas MB-EE with a small radius ($\theta = 0.05$) does best when the errors are heteroskedastic. Moreover, when the errors are heteroskedastic and the model is under-fitted, MB has a higher RMSE than HI. Consequently, when the researcher is focusing on the ATT and believes that conditional independence holds, but is unsure of the proper functional form, MB-EE with a small radius is recommended as a sensitivity check as it performs significantly better than HI and MB when errors are heteroskedastic, and only marginally worse than MB (but still much better than HI) when errors are homoskedastic.

Finally, consider the case of selection on unobservables (Tables A2, A3, A5, and A6). Upon digesting the results, two salient patterns emerge. First, as in Tables A1 and A4, when under-fitting the model,

MB and MB-EE with a small radius ($\theta = 0.05$) perform much better than the remaining estimators in most cases, with MB doing marginally better for the ATE under heteroskedasticity and the ATT under homoskedasticity. Second, when using the correct specification or when over-fitting, KV, BVN, and MB-BC with a large radius ($\theta = 0.25$) are comparable overall and generally outperform the remaining estimators. However, the relative ranking among these three varies depending on the parameter of interest, whether the errors are normal and/or homoskedastic, and whether the model is correctly specified or over-specified.

For example, with normal errors (Panel I in each of the tables), BVN and MB-BC have similar RMSEs and outperform KV although the margin is narrower (or eliminated) when the model is correctly specified and the errors are heteroskedastic; unlike most of the other estimators, the performance of KV is harmed by over-fitting. This same pattern tends to hold with non-normal errors in Tables A2 and A5 (modest selection on unobservables). However, in Panel II of Table A3 (constant treatment effect with non-normal errors), KV tends to do very well, particularly when the errors are heteroskedastic or the focus is on the ATT. Thus, with relatively strong selection on unobservables and non-normal errors, KV is superior. In contrast, in Panel II of Table A6 (heterogeneous treatment effect with non-normal errors), BVN does quite well when the focus is on the ATE (with MB-BC with a large radius ($\theta = 0.25$) performing only marginally worse) or the ATT with homoskedastic errors, whereas KV does best when estimating the ATT with heteroskedastic errors.

5.2.3 Discussion

In light of the Monte Carlo results, the applied researcher is presented with a new set of tools that should prove useful in assessing the causal impact of a treatment. If one believes that individuals select into the treatment on the basis of observables only, but the proper functional form is unknown, then it is best to use HI and definitely err on the side of over-fitting the propensity score model. Moreover, our MB or MB-EE estimator with a large radius provides a useful robustness check to HI as it performs nearly as well when the model is correctly specified or over-specified, and better when the propensity score model is under-specifed.

If one believes that individuals select into the treatment on the basis of unobservables, but the proper functional form is unknown, then KV, BVN, and MB-BC perform well in general, with KV performing marginally better (worse) when the error in the treatment equation is heteroskedastic (homoskedastic). In addition, despite the heavy reliance of the MB-BC estimator on the BVN selection model, there are multiple instances where MB-BC has a lower RMSE than BVN. Finally, it is again best to err on the side of over-fitting the model, and there is no advantage gained by using the BVN or MB-BC estimators allowing for non-normality.

6 Application

6.1 Background

As recognition of the importance of breakfast to overall health, as well as academic achievement, has grown, so to has the popularity and scrutiny of the SBP. The SBP is a federally funded program, overseen by the US Department of Agriculture (USDA), but administered by state education agencies. The SBP was established in 1966 by the Child Nutrition Act, and made permanent in 1975. Participation by schools – both public and private – is voluntary (unless mandated by the state). In 1970, roughly 0.5 million students were served on an average school day. This figure increased to 4.0 million in 1990, 7.5 million in 2000, and 10.5 million in 2008 (8.1 million of which were free or reduced-price meals).[15] The program cost the federal government $2.4 billion in 2008.

If schools do participate, they are reimbursed a fixed amount per breakfast served.[16] However, to qualify for reimbursement, the meals must meet federal nutrition guidelines established in 1995 under the "School Meals Initiative for Healthy Children" (SMI). This entails no more than 30% of the breakfast's calories be derived from fat, and less than 10% from saturated fat. Breakfasts also must provide one-fourth of the Recommended Dietary Allowance for protein, calcium, iron, Vitamin A, Vitamin C, and contain an age-appropriate level of calories.

Enforcement of the SMI requirements is handled by requiring states to monitor local school food authorities through reviews conducted at least once every five years. In turn, the USDA monitors state compliance with this review requirement. The USDA has also begun to provide regional and local training to ensure adequate overview, as well as training in the preparation of healthy meals and dissemination of information for children related to the importance of a healthy diet.

6.2 Data

To analyze the impact of SBP participation on child health, we utilize data from the *Early Childhood Longitudinal Study-Kindergarten Class of 1998-99* (ECLS-K). We measure participation in the SBP in spring first grade; we ignore kindergarten participation due to many children attending half-day programs. Our outcomes of interest are measures of child health in spring third and fifth grade or the change from fall first grade to spring third and fifth grade. As such, we are analyzing more of the long-run relationship

[15] 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. In addition, children from households that receive aid through food stamps, Temporary Assistance for Needy Families, or the Food Distribution Program on Indian Reservations are automatically eligible for free meals.

[16] For the 2009-2010 school year, reimbursement rates are $1.46 per free meal, $1.16 per reduced-price breakfast, and $0.26 per full-priced breakfast. Schools establish their own prices for full price meals, but prices for reduced price meals are capped.

between child health and SBP participation. We utilize three measures of child health:

(i) growth rate in BMI (i.e., change in log BMI) from first grade to spring third and fifth grade,

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

(iii) indicator for obesity status in spring third and fifth grade,

where we define overweight (obesity) as having a BMI above the (85th) 95th percentile.[17]

The following covariates are included in X: child's race (white, black, Hispanic, and other), age, and gender, child's birth weight, household socioeconomic (SES) status, 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 prior to kindergarten, region, city type (urban, suburban, or rural), a measure of food security, and whether the child eats a school-provided lunch.

Children with missing data for gender and age are dropped from our sample. We also restrict the sample to public schools. Missing values for the remaining control variables are imputed and imputation dummies are added to the control set. In addition, we restrict attention to public schools. The final sample contains 9,952 students when analyzing third grade outcomes, of which 3,071 participate in the SBP. The sample size falls to 7,824 when we analyze fifth grade outcomes. Table 3 provides summary statistics.

In light of the Monte Carlo results, we start with a linear model and then gradually add higher order and interaction terms for the continuous covariates as a specification check. In addition, in the interest of brevity, we do not present any of the estimators based on non-normality given their poor performance; even if the model is under-specified, MB is preferred to MB-EE when the errors are heteroskedastic (as they are in this case) and the focus is on the ATE.[18]

6.3 Results

6.3.1 Baseline Specifications

The results are presented in Tables 4–6, where Table 4 contains the results using BMI growth to measure child health and Table 5 (Table 6) displays the results using overweight (obesity) status as the outcome. Panel A in each table displays the estimates of the ATE, while Panel B presents the ATT estimates. As stated above, we present the results from three specifications: Specification (1) includes only linear terms in X; Specification (2) adds squared and interaction terms for the continuous covariates; Specification (3)

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

[18] While we present results for both the ATE and ATT, the ATE is of more relevance to policymakers interested in expanding participation in the SBP to reach new students in order to potentially combat the rise in childhood obesity.

includes cubic terms for the continuous covariates and interaction terms among the linear and squared terms involving the continuous covariates. Finally, 90% confidence intervals for all the estimates are obtained by bootstrap with 250 repetitions.[19]

Turning to the results in Table 4, OLS and HI indicate a positive and statistically significant association between SBP and BMI growth regardless of whether one measures growth between first and third grade or first and fifth grade, regardless of whether one focuses on the ATE or ATT, and regardless of model specification. In terms of magnitudes, the results indicate that SBP participation in first grade is associated with roughly a 1% (2.5%) *increase* in BMI growth for the average student between first and third (fifth) grade according to HI.

If we consider the model to be under-specified, then the MB estimates with a small radius ($\theta = 0.05$) are preferred and fail to find a statistically meaningful relationship between SBP and BMI growth. Notably, while this arises in part due to a loss in precision, it also reflects in part a reduction in the point estimates. Relative to the ATE, the point estimates fall in magnitude between eight and 72 percent. Relative to ATT, the point estimates fall between 75 and 111 percent using the third grade outcome, while they remain virtually constant using the fifth grade outcome.

When we move on to the estimators designed to circumvent selection on unobservables, the point estimates become negative regardless of outcome, parameter of interest, or specification. Erring on the side of over-specifying the model, the KV, BVN, and MB-BC estimators each roughly indicate a 1% *decrease* in BMI growth for the average student between first and third grade according to Specification (3).[20] In addition, the estimates are statistically significant in all cases except for KV. For the fifth grade outcome, the estimates range from a 2.1% (KV) to 4.4% (BVN) decline in BMI growth for the average student, although none are statistically significant. In terms of the ATT estimates, the point estimates range from 1.2% to 3.7% depending on the estimator and the outcome. Again, none of the estimates are statistically significant.

In sum, the results in Table 4 indicate the importance of addressing positive selection into SBP on the basis of weight trajectories in order to estimate the causal effect of participation, as suggested in MTH. Moreover, consistent with this view, the MB estimator is able to reduce the selection bias sufficiently such that the positive and statistically significant effect obtained by estimators that require conditional independence disappears. Finally, the KV, BVN, and MB-BC estimators each appear able to circumvent this non-random selection and indicate, at worst, no impact of SBP on child weight and, at best, a small,

[19] Note, when obtaining confidence intervals for MB and MB-BC, we re-estimate $\widehat{P^*(X)}$ within each bootstrap repetition.

[20] For the KV estimator, we note that for *all* models estimated in this study, we reject the null of homoskedastic errors at the $p < 0.01$ confidence level. In addition, the F-statistic on the instrument in the first-stage is always well in excess of Stock et al.'s (2002) rule-of-thumb value of ten. Detailed results are available from the authors upon request.

beneficial causal effect of participation.

Tables 5 and 6 present the results using overweight and obesity status as the outcome of interest. Qualitatively, the pattern of results across estimators, specifications, and parameters of interest are unchanged from Table 4. The magnitudes, however, do suggest an economically more meaningful impact of SBP participation. Specifically, according to the preferred Specification (3), the HI estimates indicate that SBP participation is associated with a 4.8% (3.4%) *increase* in the probability of being overweight in the third grade for the average student (average student in the treatment group); a 2.8% (3.2%) *increase* in the probability of being obese in the third grade for the average student (average student in the treatment group). The point estimates are even larger when we measuring child status in fifth grade. However, the corresponding estimates KV, BVN, and MB-BC estimates suggest a 8.8 to 23.0% decline in the probability of being overweight in the third grade depending on which estimator one uses and whether one focuses on the ATE or ATT; 10.8 to 27.8% when using fifth grade overweight status. Similarly, the three estimators indicate a 7.6 to 18.2% decline in the probability of being obese in the third grade; 4.3 to 13.4% when using fifth grade obesity status. As in Table 4, the KV estimates are never statistically significant; the BVN and MB-BC estimates are in some instances.

The combined results do paint a picture consistent with the results in MTH. Specifically, without controlling for any selection on unobservables, SBP appears to be a contributing factor to the childhood obesity crisis. However, after controlling for some of this selection by minimizing the bias, the results are weakened somewhat. Finally, when we remove the bias entirely using the KV, BVN, or MB-BC estimator, there is no longer any evidence that SBP contributes to childhood obesity, and in fact there is some statistically meaningful evidence to the contrary.

6.3.2 Further Investigations

Before concluding, we extend our analysis in two directions. The first direction relates to the presence of underweight children. When conceived, the original focus of the SBP (and the NSLP) was on providing all children with a minimum level of nutrition (Guthrie et al. 2009). Only recently, with the rise in childhood obesity, has concern shifted to the upper end of the weight distribution, with these programs trying to help children maintain a healthy weight (as opposed to being overweight). However, it is important to examine if the SBP has lost site of its original mission in bringing underweight children into a healthy weight range. To that end, we create a binary indicator for underweight status, defined as one if children have a BMI below the 20th percentile.[21] In light of the positive selection into the SBP on the basis of weight discussed

[21] The CDC defines underweight as being below the fifth percentile. However, there are very few children below the fifth percentile in the data; thus, we use a higher cut-off. In terms of trends, using the official CDC definition of underweight, the incidence of underweight children between the ages of six and eleven has declined from 5.3% in 1971–1974 to 2.7% in

previously, now we expect the selection on observables estimators to be biased downward.

The results are presented in Table 7. Consistent with our expectations, OLS and HI indicate a negative association between SBP and the probability of being underweight regardless of whether one measures weight in third or fifth grade, regardless of whether one focuses on the ATE or ATT, and regardless of model specification. In nearly all cases, the estimate is statistically significant as well. In terms of magnitudes, the results indicate that SBP participation in first grade is associated with roughly a 2.5% decrease in the probability of being underweight in both third and fifth grade for the average student according to HI.

If we consider the model to be under-specified, then the MB estimates with a small radius ($\theta = 0.05$) are preferred and fail to find a statistically meaningful relationship between SBP and being underweight. This arises, by and large, from a decline in precision although the point estimates for the ATE do become positive in Specifications (1) and (2) for fifth grade. Finally, when we turn to the estimators designed to circumvent selection on unobservables, the point estimates are either negative and very close to zero or become positive. While this is consistent the positive selection discussed in Tables 4–6, none of the estimates are statistically significant. Interpreting these statistically insignificant estimates as evidence of no causal effect of SBP participation on the probability of being underweight, in combination with our prior findings in Tables 4–6, indicates that SBP combats childhood obesity without sacrificing its original goal of ensuring children are not malnourished.

Our final investigation relates to the possibility of measurement error in reported program participation. To mitigate the possibility of mis-measurement, we follow a strategy loosely based on Black et al. (2000) who note that one can reduce the bias from measurement error in a binary regressor if one has two mis-measured indicators and the measurement errors are independent. In such a case, an improved estimate is obtained by defining a binary variable equal to one if both mis-measured indicators are equal to one. To that end, we re-estimate the models in Tables 4–7 except now we define the treatment as one if the student reported eating breakfast at school in first and third grade (when using third grade outcomes) and first, third, and fifth grade (when using fifth grade outcomes).[22]

In the interest of brevity, the results are relegated to Appendix C and we simply note that the results are qualitatively unchanged. The two most noteworthy differences are the fact that the KV estimates are now statistically significant in many instances, and the fact that the estimated effects on fifth grade outcomes are statistically significant more often. This latter result presumably is due more to the new

2003–2006 (http://www.cdc.gov/nchs/data/hestat/underweight_children.pdf).

[22] This exercise is merely meant to be suggestive. First, the assumption that measurement error in participation reports across waves are independent is open for debate. Second, unlike Black et al. (2000), our multiple indicators come from different points in time. As such, if there is no measurement error in the data, we are implicitly re-defining the treatment relative to the treatment analyzed in Tables 4–7.

definition of the treatment rather than a correction for measurement error.

7 Conclusion

The program evaluation literature has expanded rapidly over the past decade. While our knowledge concerning methods that are designed to provide consistent estimates of some measure of the causal effect of a binary treatment under conditional independence, as well as typical IV methods when conditional independence fails, is relatively well developed, researchers are less informed about how to proceed when conditional independence fails yet the usual type of exclusion restrictions is unavailable. In this study, we aim to reduce this deficit by proposing two new estimators for this situation, and then evaluating the performance of our estimators as well as that of bivariate normal selection model, the selection model extended to the case of non-normality, and a recently proposed IV estimator proposed in Klein and Vella (2009) that relies on heteroskedasticity. In addition, we use these estimators to assess the causal impact of a program of great interest to policymakers: the School Breakfast Program.

Our analysis leads to some general guidelines that applied researchers may wish to follow in similar situations moving forward. First, in applications where the researcher believes conditional independence holds, our minimum-biased estimator offers a nice robustness check since it performs nearly as well when the model is correctly specified or over-specified, but vastly better when the model is under-specified. Second, when conditional independence does not hold but the model is correctly specified or over-specified, our bias-corrected estimator does quite well, even outperforming estimators relying purely on functional form for identification. Third, our parametric Klein and Vella (2009) estimator performs very well when the error in the treatment equation is in fact heteroskedastic, but still does well when this is not the case. Finally, consistent with Millimet and Tchernis (2009), in all cases the penalty to over-specifying the model, if there is one at all, pales in comparison to the penalty from under-specifying the model. However, the KV estimator appears to be most sensitive to over-fitting.

In terms of our analysis of the SBP, the various estimators offer a coherent picture of the causal effect of the program. Specifically, we find a positive and statistically significant association between SBP and child weight when using estimators that require conditional independence. The association remains positive, but becomes statistically insignificant, when we use our minimum-biased estimator. Finally, consistent with the suggestive evidence in MTH, as well as Bhattacharya et al. (2006), we find a negative and mostly statistically significant causal effect of SBP participation on child weight using the bivariate normal selection model and Klein and Vella's (2009) estimator, as well as our bias-corrected estimator.

While promising, further analysis on the econometric side may prove fruitful by considering the choice

of θ or differential weighting of observations within the neighborhood around P^*. In addition, one may wish to consider a bias-corrected estimator of the *unconditional* average treatment effect when the treatment effect is heterogeneous. For instance, one might wish to obtain a bias-corrected estimate conditional on the propensity score, $P(X)$, and then estimate the unconditional average treatment effect by taking the expectation of this over the distribution of X in the population (or sub-population of treated). Finally, future work may compare the estimators considered here to other estimators that circumvent the need for an exclusion restriction; for instance, those considered in Ebbes et al. (2009) for continuous endogenous regressors. In terms of the analysis of the SBP, future analysis into the impact of participation on other outcomes, such as cognitive and non-cognitive skills, will provide policymakers with better information about the potential benefits of the program.

References

[1] Abadie, A. and G.W. Imbens (2009), "Matching on the Estimated Propensity Score," unpublished manuscript, Harvard University.

[2] Altonji, J.G., T.E. Elder, and C.R. Taber (2005), "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113, 151-184.

[3] 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.

[4] Black, D.A., M.C. Berger, and F.A. Scott (2000), "Bounding Parameter Estimates with Nonclassical Measurement Error," *Journal of the American Statistical Association*, 95, 739-748

[5] 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.

[6] Bleich, S., D. Cutler, C. Murray, and A. Adams (2007), "Why Is The Developed World Obese?" NBER Working Paper No. 12954.

[7] 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.

[8] Dehejia, R. H., and S. Wahba (1999), "Casual Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053-1062.

[9] Ebbes, P., M. Wedel, and U. Böckenholt (2009), "Frugal IV Alternatives to Identify the Parameter for an Endogenous Regressor," *Journal of Applied Econometrics*, 24, 446-468.

[10] Finkelstein, E.A., J.G. Trogdon, J.W. Cohen, and W. Dietz (2009), "Annual Medical Spending Attributable to Obesity: Payer- and Service-Specific Estimates," *Health Affairs*, 28, w822-w831.

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

[12] Guthrie, J., C. Newman, and K. Ralston (2009), "USDA School Meal Programs Face New Challenges," *Choices*, 24, http://www.choicesmagazine.org/magazine/article.php?article=83.

[13] Heckman, J. and S. Navarro-Lozano (2004), "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models," *Review of Economics and Statistics*, 86, 30-57

[14] Heckman, J.J. and R. Robb (1985), "Alternative Methods for Evaluating the Impact of Interventions," in J.J. Heckman and B. Singer (eds.) *Longitudinal Analysis of Labor Market Data*, Cambridge, England: Cambridge University Press.

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

[16] Headrick, T.C. and S.S. Sawilowsky (1999), "Simulating Correlated Multivariate Nonnormal Distributions: Extending the Fleishman Power Method," *Psychometrika*, 64, 25-35.

[17] 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.

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

[19] 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.

[20] Imbens, G.W. (2004), "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *Review of Economics and Statistics*, 86, 4-29.

[21] Imbens, G., and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-475.

[22] Imbens, G.W. and J.M. Wooldridge (2009), "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47, 5-86.

[23] Klein, R. and F. Vella (2009), "A Semiparametric Model for Binary Response and Continuous Outcomes Under Index Heteroskedasticity," *Journal of Applied Econometrics*, 24, 735-762.

[24] Lee, L.-F. (1984), "Tests for the Bivariate Normal Distribution in Econometric Models with Selectivity," *Econometrica*, 52, 843-863.

[25] Lee, D.S. and T. Lemieux (2009), "Regression Discontinuity Designs in Economics," NBER WP No. 14723.

[26] Mardia, K.V. (1970), *Families of Bivariate Distributions*, Darien, Connecticut: Hafner Publishing Company.

[27] Millimet, D.L. and R. Tchernis (2009), "On the Specification of Propensity Scores: with Applications to the Analysis of Trade Policies," *Journal of Business & Economic Statistics*, 27, 397-415.

[28] Millimet, D.L., R. Tchernis, and M. Hussain (2009), "School Nutrition Programs and the Incidence of Childhood Obesity," *Journal of Human Resources*, forthcoming.

[29] 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).

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

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

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

[33] 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.

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

[35] Schanzenbach, D.W. (2008), "Does the Federal School Lunch Program Contribute to Childhood Obesity?" *Journal of Human Resources*, forthcoming.

[36] Smith, J.A. and P.E. Todd (2005), "Does Matching Overcome LaLonde's Critique?" *Journal of Econometrics*, 125, 305-353.

[37] Stock, J.H., J.H. Wright, and M. Yogo (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518-529.

[38] Trasande, L., Y. Liu, G. Fryer, and M. Weitzman (2009), "Effects of Childhood Obesity on Hospital Care and Costs, 1999-2005," *Health Affairs*, 28, w751-w760.

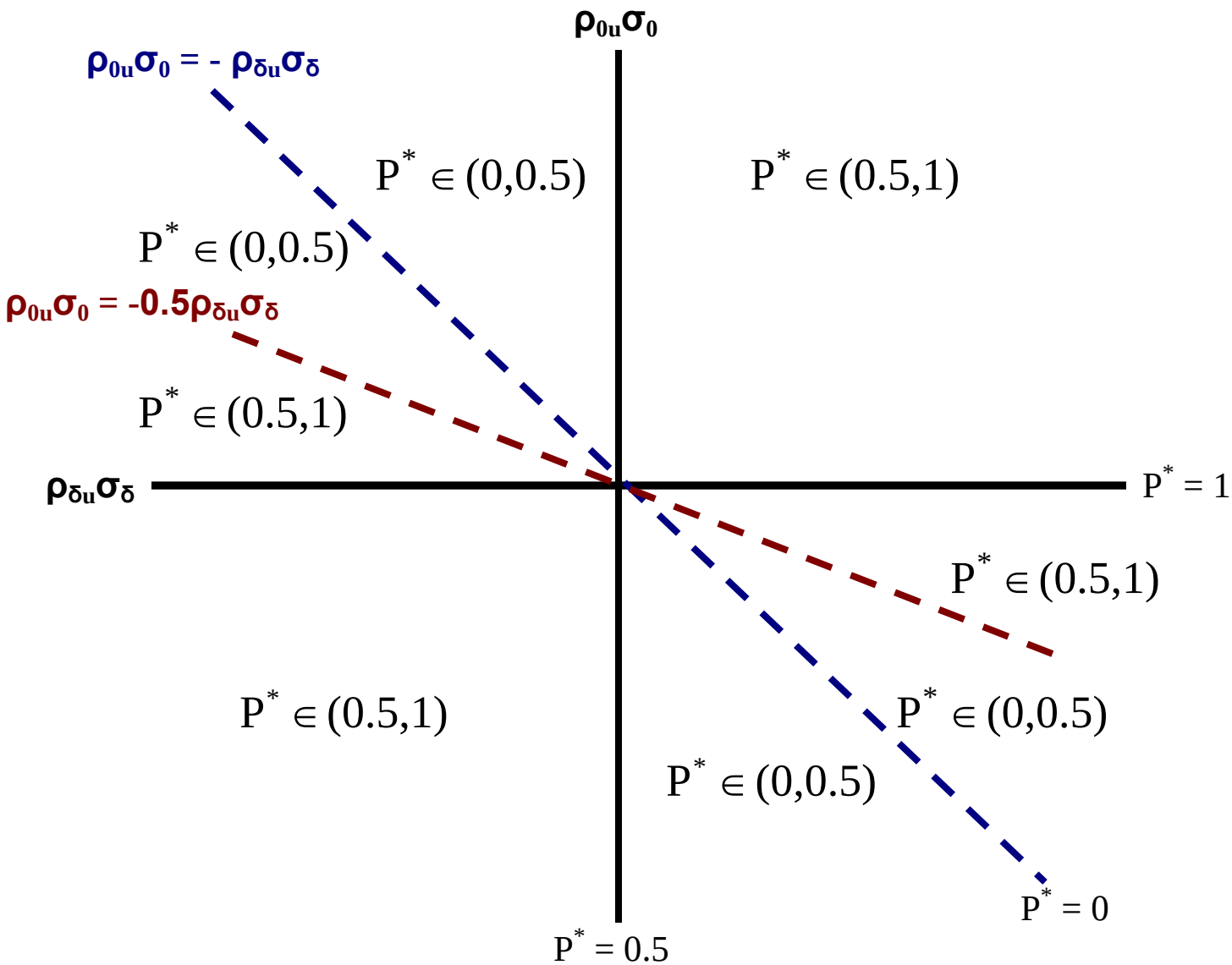

Figure 1. Bias-Minimizing Value of the Propensity Score Under Different Parameter Values.

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

Table 1. Monte Carlo Results: Estimates in the Common Effect Model ($\tau_i = 1$)

	Homoskedastic Error in Treatment Equation						Heteroskedastic Error in Treatment Equation					
	ATE			ATT			ATE			ATT		
	$\rho_0\sigma_0 = 0$	$\rho_0\sigma_0 = -0.25$	$\rho_0\sigma_0 = -0.50$	$\rho_0\sigma_0 = 0$	$\rho_0\sigma_0 = -0.25$	$\rho_0\sigma_0 = -0.50$	$\rho_0\sigma_0 = 0$	$\rho_0\sigma_0 = -0.25$	$\rho_0\sigma_0 = -0.50$	$\rho_0\sigma_0 = 0$	$\rho_0\sigma_0 = -0.25$	$\rho_0\sigma_0 = -0.50$
	$\rho_{01} = 1$	$\rho_{01} = 1$	$\rho_{01} = 1$	$\rho_{01} = 1$	$\rho_{01} = 1$	$\rho_{01} = 1$	$\rho_{01} = 1$	$\rho_{01} = 1$	$\rho_{01} = 1$	$\rho_{01} = 1$	$\rho_{01} = 1$	$\rho_{01} = 1$
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
I. Normally Distributed Errors												
τ_{HI}	0.047	0.440	0.883	0.053	0.443	0.893	0.047	0.415	0.841	0.048	0.421	0.844
$\tau_{MB,0.05}$	0.113	0.407	0.795	0.102	0.411	0.800	0.138	0.406	0.817	0.114	0.432	0.817
$\tau_{MB,0.25}$	0.064	0.402	0.805	0.056	0.404	0.809	0.060	0.410	0.823	0.060	0.428	0.831
$\tau_{MB,EE,0.05}$	0.129	0.471	0.892	0.123	0.464	0.898	0.176	0.388	0.788	0.169	0.420	0.819
$\tau_{MB,EE,0.25}$	0.068	0.437	0.860	0.067	0.437	0.864	0.071	0.411	0.829	0.070	0.421	0.836
τ_{KV}	0.360	0.374	0.368	0.360	0.374	0.368	0.238	0.279	0.269	0.238	0.279	0.269
τ_{BVN}	0.301	0.287	0.278	0.226	0.245	0.312	0.232	0.265	0.300	0.177	0.259	0.394
$\tau_{MB\text{-}BC,0.05}$	0.312	0.301	0.287	0.321	0.296	0.294	0.257	0.266	0.317	0.259	0.294	0.334
$\tau_{MB\text{-}BC,0.25}$	0.296	0.291	0.285	0.307	0.283	0.277	0.225	0.264	0.314	0.240	0.278	0.328
$\tau_{BVN,\ EE}$	1.610	1.363	1.402	1.068	0.933	0.964	0.828	0.903	0.835	0.548	0.617	0.585
$\tau_{MB\text{-}BC,EE,0.05}$	1.319	1.268	1.422	1.381	1.185	1.247	0.730	0.823	0.791	0.688	0.761	0.731
$\tau_{MB\text{-}BC,EE,0.25}$	1.289	1.244	1.396	1.348	1.171	1.223	0.698	0.803	0.781	0.673	0.746	0.733
II. Asymmetric, Non-Normally Distributed Errors												
τ_{HI}	0.050	0.378	0.628	0.055	0.348	0.489	0.048	0.364	0.641	0.051	0.339	0.511
$\tau_{MB,0.05}$	0.097	0.350	0.391	0.100	0.420	0.784	0.125	0.338	0.348	0.112	0.439	0.806
$\tau_{MB,0.25}$	0.059	0.348	0.449	0.055	0.387	0.715	0.064	0.346	0.446	0.059	0.409	0.731
$\tau_{MB,EE,0.05}$	0.130	0.434	0.792	0.128	0.348	0.518	0.171	0.344	0.616	0.161	0.325	0.487
$\tau_{MB,EE,0.25}$	0.069	0.385	0.648	0.069	0.338	0.496	0.069	0.354	0.571	0.067	0.329	0.492
τ_{KV}	0.363	0.400	0.368	0.363	0.400	0.368	0.248	0.275	0.257	0.248	0.275	0.257
τ_{BVN}	0.281	0.308	0.380	0.207	0.300	0.585	0.245	0.288	0.364	0.184	0.300	0.572
$\tau_{MB\text{-}BC,0.05}$	0.286	0.313	0.378	0.289	0.381	0.868	0.273	0.294	0.373	0.257	0.395	0.866
$\tau_{MB\text{-}BC,0.25}$	0.277	0.303	0.396	0.278	0.350	0.801	0.242	0.285	0.390	0.248	0.363	0.794
$\tau_{BVN,\ EE}$	1.608	1.631	1.819	1.071	1.083	1.172	0.899	0.900	1.080	0.598	0.590	0.700
$\tau_{MB\text{-}BC,EE,0.05}$	1.329	1.514	1.764	1.401	1.389	1.514	0.838	0.903	1.169	0.787	0.774	0.956
$\tau_{MB\text{-}BC,EE,0.25}$	1.296	1.503	1.786	1.377	1.374	1.443	0.815	0.872	1.125	0.762	0.749	0.874

Notes: Numbers reflect the root mean squared error based on 250 simulated data sets with 5,000 observations. HI = Hirano and Imbens (2001) estimator; MB = minimum-biased estimator using a cut-off level (α) chosen to retain 5% or 25% of the treatment and control groups; KV = Klein and Vella (2009) estimator; BVN = Heckman bivariate normal selection model; MB-BC = bias corrected estimator using a cut-off level (α) chosen to retain 5% or 25% of the treatment and control groups. EE implies estimator relies on the Edgeworth expansion formula. Shading indicates best performance within each column. See text for further details.

Table 2. Monte Carlo Results: Heterogeneous Effect Model ($\tau_i = 1 + \delta_i$)

	Homoskedastic Error in Treatment Equation						Heteroskedastic Error in Treatment Equation					
	ATE			ATT			ATE			ATT		
	$\rho_0\sigma_0 = 0$	$\rho_0\sigma_0 = -0.20$	$\rho_0\sigma_0 = -0.40$	$\rho_0\sigma_0 = 0$	$\rho_0\sigma_0 = -0.20$	$\rho_0\sigma_0 = -0.40$	$\rho_0\sigma_0 = 0$	$\rho_0\sigma_0 = -0.20$	$\rho_0\sigma_0 = -0.40$	$\rho_0\sigma_0 = 0$	$\rho_0\sigma_0 = -0.20$	$\rho_0\sigma_0 = -0.40$
	$\rho_{01} = 0.50$	$\rho_{01} = 0.50$	$\rho_{01} = 0.50$	$\rho_{01} = 0.50$	$\rho_{01} = 0.50$	$\rho_{01} = 0.50$	$\rho_{01} = 0.50$	$\rho_{01} = 0.50$	$\rho_{01} = 0.50$	$\rho_{01} = 0.50$	$\rho_{01} = 0.50$	$\rho_{01} = 0.50$
	$\rho_\delta\sigma_\delta = 0$	$\rho_\delta\sigma_\delta = -0.10$	$\rho_\delta\sigma_\delta = -0.10$	$\rho_\delta\sigma_\delta = 0$	$\rho_\delta\sigma_\delta = -0.10$	$\rho_\delta\sigma_\delta = -0.10$	$\rho_\delta\sigma_\delta = 0$	$\rho_\delta\sigma_\delta = -0.10$	$\rho_\delta\sigma_\delta = -0.10$	$\rho_\delta\sigma_\delta = 0$	$\rho_\delta\sigma_\delta = -0.10$	$\rho_\delta\sigma_\delta = -0.10$
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
I. Normally Distributed Errors												
τ_{HI}	0.046	0.412	0.768	0.052	0.358	0.719	0.046	0.390	0.734	0.046	0.338	0.683
$\tau_{MB,0.05}$	0.100	0.379	0.713	0.108	0.326	0.653	0.125	0.385	0.730	0.115	0.352	0.680
$\tau_{MB,0.25}$	0.062	0.378	0.709	0.056	0.308	0.645	0.062	0.378	0.725	0.061	0.327	0.662
$\tau_{MB,EE,0.05}$	0.124	0.463	0.810	0.111	0.333	0.697	0.158	0.368	0.693	0.169	0.293	0.640
$\tau_{MB,EE,0.25}$	0.066	0.413	0.756	0.065	0.304	0.671	0.068	0.380	0.719	0.071	0.290	0.644
τ_{KV}	0.361	0.369	0.369	0.361	0.370	0.373	0.258	0.254	0.268	0.258	0.262	0.275
τ_{BVN}	0.288	0.262	0.261	0.213	0.220	0.266	0.253	0.266	0.284	0.191	0.238	0.325
$\tau_{MB\text{-}BC,0.05}$	0.299	0.271	0.264	0.298	0.274	0.271	0.276	0.262	0.302	0.283	0.297	0.321
$\tau_{MB\text{-}BC,0.25}$	0.284	0.259	0.264	0.284	0.263	0.263	0.246	0.253	0.289	0.252	0.275	0.299
$\tau_{BVN,\ EE}$	1.585	1.708	1.434	1.053	1.145	0.977	0.867	0.868	0.852	0.582	0.575	0.586
$\tau_{MB\text{-}BC,EE,0.05}$	1.389	1.431	1.375	1.383	1.471	1.243	0.794	0.942	0.973	0.731	0.743	0.802
$\tau_{MB\text{-}BC,EE,0.25}$	1.360	1.401	1.362	1.357	1.450	1.214	0.767	0.916	0.962	0.708	0.737	0.789
II. Asymmetric, Non-Normally Distributed Errors												
τ_{HI}	0.045	0.380	0.604	0.051	0.287	0.443	0.044	0.366	0.607	0.046	0.287	0.465
$\tau_{MB,0.05}$	0.111	0.304	0.325	0.102	0.314	0.695	0.127	0.269	0.238	0.106	0.399	0.727
$\tau_{MB,0.25}$	0.056	0.311	0.401	0.051	0.272	0.632	0.063	0.305	0.400	0.060	0.362	0.662
$\tau_{MB,EE,0.05}$	0.124	0.496	0.774	0.120	0.242	0.443	0.159	0.417	0.609	0.151	0.294	0.470
$\tau_{MB,EE,0.25}$	0.067	0.397	0.621	0.066	0.210	0.440	0.065	0.368	0.554	0.068	0.275	0.465
τ_{KV}	0.400	0.380	0.437	0.400	0.391	0.454	0.273	0.270	0.316	0.273	0.279	0.328
τ_{BVN}	0.298	0.279	0.289	0.221	0.215	0.322	0.254	0.252	0.250	0.192	0.245	0.382
$\tau_{MB\text{-}BC,0.05}$	0.310	0.291	0.329	0.305	0.281	0.520	0.281	0.285	0.303	0.280	0.342	0.623
$\tau_{MB\text{-}BC,0.25}$	0.291	0.280	0.300	0.295	0.256	0.464	0.250	0.254	0.278	0.260	0.317	0.561
$\tau_{BVN,\ EE}$	1.405	1.641	1.772	0.931	1.075	1.155	0.861	0.953	0.915	0.566	0.630	0.631
$\tau_{MB\text{-}BC,EE,0.05}$	1.235	1.472	1.797	1.248	1.377	1.486	0.831	0.992	1.031	0.754	0.790	0.827
$\tau_{MB\text{-}BC,EE,0.25}$	1.206	1.484	1.860	1.221	1.369	1.456	0.810	0.976	1.014	0.733	0.771	0.779

Notes: See Table 1.

Table 3. 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)	0.309	0.462	1	0	0	0
Third Grade Child Weight						
BMI Growth Rate	0.104	0.090	0.116	0.095	0.099	0.087
Overweight (1 = Yes)	0.365	0.482	0.421	0.494	0.341	0.474
Obese (1 = Yes)	0.197	0.398	0.238	0.426	0.179	0.383
Fifth Grade Child Weight						
BMI Growth Rate	0.198	0.122	0.222	0.123	0.188	0.119
Overweight (1 = Yes)	0.414	0.493	0.492	0.500	0.380	0.485
Obese (1 = Yes)	0.230	0.421	0.293	0.455	0.202	0.402
Controls						
Age (in months)	109.448	4.334	109.578	4.468	109.390	4.271
Gender (1 = boy)	0.513	0.500	0.508	0.500	0.515	0.500
White (1 = Yes)	0.554	0.497	0.344	0.475	0.648	0.478
Black (1 = Yes)	0.137	0.344	0.269	0.444	0.078	0.268
Hispanic (1 = Yes)	0.191	0.393	0.265	0.442	0.157	0.364
Child's Birthweight (ounces)	117.066	26.044	115.256	27.249	117.873	25.449
Child's Birthweight (1 = Missing)	0.068	0.252	0.106	0.309	0.051	0.220
Central City (1 = Yes)	0.328	0.470	0.387	0.487	0.302	0.459
Urban Fringe & Large Town (1 = Yes)	0.388	0.487	0.260	0.438	0.445	0.497
Northeast (1 = Yes)	0.175	0.380	0.110	0.313	0.203	0.403
Midwest (1 = Yes)	0.239	0.426	0.188	0.391	0.261	0.439
South (1 = Yes)	0.338	0.473	0.464	0.499	0.281	0.450
Mother's Age at First Birth (Years)	23.567	5.090	21.290	4.232	24.584	5.111
Mother's Age at First Birth (1 = Missing)	0.117	0.321	0.164	0.370	0.095	0.294
WIC Benefits Prior to Kindergarten (1 = Yes)	0.460	0.498	0.733	0.443	0.338	0.473
WIC Benefits Prior to Kindergarten (1 = Missing)	0.031	0.173	0.046	0.209	0.024	0.154
Mother's Education = Less Than High School (1 = Yes)	0.148	0.355	0.275	0.446	0.091	0.288
Mother's Education = High School (1 = Yes)	0.312	0.463	0.376	0.484	0.284	0.451
Mother's Education = Some College (1 = Yes)	0.307	0.461	0.243	0.429	0.336	0.472
Mother's Education = Bachelor's Degree (1 = Yes)	0.136	0.343	0.045	0.206	0.177	0.382
Mother's Education = Advanced College Degree (1 = Yes)	0.065	0.247	0.016	0.125	0.087	0.282
Mother Employed Full-Time During Kindergarten (1 = Yes)	0.403	0.491	0.401	0.490	0.404	0.491
Mother Employed Part-Time During Kindergarten (1 = Yes)	0.189	0.391	0.131	0.337	0.214	0.410
No Mother in household During Kindergarten (1 = Yes)	0.288	0.453	0.309	0.462	0.278	0.448
SES Index	-0.055	0.771	-0.490	0.678	0.139	0.729
SES Index (1 = Missing)	0.016	0.126	0.026	0.160	0.012	0.107
Never Worried About Running Out of Food in Household (1 = Yes)	0.793	0.405	0.687	0.464	0.841	0.366
Never Worried About Running Out of Food in Household (1 = Missing)	0.100	0.300	0.136	0.343	0.084	0.277
NSLP Participation (1 = Yes)	0.543	0.498	0.755	0.430	0.448	0.497
NSLP Participation (1 = Missing)	0.049	0.215	0.066	0.249	0.041	0.197
Number of Children's Books in Household	72.250	56.149	51.441	48.070	81.537	57.003
Number of Children's Books in Household (1 = Missing)	0.109	0.311	0.143	0.350	0.093	0.291

Notes: N = 9,952 (full sample for third grade outcomes); of this, 3,071 are SBP participants and 6,881 are non-participants. Data are from the ECLS-K. Change in BMI percentile and BMI growth rate calculated using baseline data from first grade.

Table 4. Effect of SBP Participation on BMI Growth

	Third Grade Outcome			Fifth Grade Outcome		
	Specification (1)	Specification (2)	Specification (3)	Specification (1)	Specification (2)	Specification (3)
A. ATE						
τ_{OLS}	0.007	0.007	0.007	0.013	0.013	0.012
	[0.003, 0.011]	[0.003, 0.011]	[0.003, 0.011]	[0.010, 0.021]	[0.011, 0.022]	[0.010, 0.021]
τ_{HI}	0.011	0.010	0.010	0.023	0.025	0.024
	[0.006, 0.016]	[0.006, 0.015]	[0.005, 0.014]	[0.016, 0.031]	[0.017, 0.030]	[0.015, 0.029]
$\tau_{MB,0.05}$	0.003	0.006	0.006	0.018	0.023	0.019
	[-0.014, 0.020]	[-0.014, 0.019]	[-0.010, 0.016]	[-0.007, 0.033]	[-0.010, 0.034]	[-0.005, 0.034]
$\tau_{MB,0.25}$	0.004	0.005	0.007	0.011	0.012	0.012
	[-0.004, 0.009]	[-0.002, 0.009]	[-0.002, 0.010]	[0.002, 0.019]	[0.003, 0.020]	[0.003, 0.021]
τ_{KV}	-0.032	-0.029	-0.029	-0.031	-0.024	-0.021
	[-0.055, -0.004]	[-0.051, 0.004]	[-0.051, 0.005]	[-0.067, 0.014]	[-0.057, 0.024]	[-0.057, 0.031]
τ_{BVN}	-0.040	-0.032	-0.032	-0.065	-0.052	-0.044
	[-0.061, -0.013]	[-0.056, -0.002]	[-0.057, -0.003]	[-0.100, -0.024]	[-0.087, -0.003]	[-0.082, 0.007]
$\tau_{MB\text{-}BC,0.05}$	-0.042	-0.031	-0.031	-0.058	-0.040	-0.036
	[-0.077, -0.003]	[-0.069, 0.007]	[-0.063, -0.001]	[-0.108, -0.012]	[-0.092, 0.004]	[-0.088, 0.015]
$\tau_{MB\text{-}BC,0.25}$	-0.040	-0.032	-0.031	-0.065	-0.051	-0.043
	[-0.068, -0.011]	[-0.060, 0.000]	[-0.061, -0.001]	[-0.101, -0.018]	[-0.088, 0.007]	[-0.082, 0.012]
P*	0.510	0.510	0.518	0.550	0.562	0.612
	[0.339, 0.815]	[0.293, 0.897]	[0.310, 0.927]	[0.403, 0.812]	[0.339, 0.919]	[0.357, 0.945]
B. ATT						
τ_{OLS}	0.007	0.007	0.007	0.013	0.013	0.012
	[0.003, 0.011]	[0.003, 0.011]	[0.003, 0.011]	[0.010, 0.021]	[0.011, 0.022]	[0.010, 0.021]
τ_{HI}	0.009	0.008	0.008	0.018	0.017	0.016
	[0.004, 0.014]	[0.004, 0.013]	[0.003, 0.013]	[0.012, 0.025]	[0.011, 0.024]	[0.010, 0.023]
$\tau_{MB,0.05}$	-0.001	0.000	0.002	0.018	0.018	0.020
	[-0.012, 0.014]	[-0.012, 0.015]	[-0.011, 0.017]	[-0.003, 0.035]	[-0.005, 0.037]	[-0.000, 0.039]
$\tau_{MB,0.25}$	0.005	0.006	0.006	0.011	0.012	0.013
	[-0.002, 0.009]	[-0.002, 0.010]	[-0.001, 0.011]	[0.005, 0.020]	[0.005, 0.023]	[0.005, 0.024]
τ_{KV}	-0.032	-0.029	-0.029	-0.031	-0.024	-0.021
	[-0.055, -0.004]	[-0.051, 0.004]	[-0.051, 0.005]	[-0.067, 0.014]	[-0.057, 0.024]	[-0.057, 0.031]
τ_{BVN}	-0.016	-0.012	-0.012	-0.022	-0.016	-0.012
	[-0.030, -0.000]	[-0.027, 0.005]	[-0.028, 0.004]	[-0.043, 0.002]	[-0.036, 0.010]	[-0.036, 0.015]
$\tau_{MB\text{-}BC,0.05}$	-0.046	-0.037	-0.034	-0.055	-0.042	-0.030
	[-0.076, -0.010]	[-0.066, 0.002]	[-0.066, 0.004]	[-0.100, -0.005]	[-0.086, 0.014]	[-0.081, 0.027]
$\tau_{MB\text{-}BC,0.25}$	-0.039	-0.031	-0.031	-0.062	-0.048	-0.037
	[-0.069, -0.011]	[-0.062, 0.001]	[-0.062, 0.001]	[-0.101, -0.012]	[-0.088, 0.013]	[-0.084, 0.017]
P*	0.500	0.500	0.500	0.500	0.500	0.500

Notes: Treatment is defined as participation in SBP in first grade. BMI growth measured from fall first grade. Specification (1) includes all variables from Table 3 entered linearly; Specification (2) adds squared and interaction terms for the continuous covariates; Specification (3) adds cubic and interaction terms for the continuous covariates. 90% empirical confidence intervals in brackets are obtained using 250 bootstrap repetitions. OLS = ordinary least squares; HI = Hirano and Imbens (2001) normalized estimator; MB = minimum-biased estimator using $\theta = 0.0.5$ or 0.25; KV = Klein and Vella (2009) estimator; BVN = Heckman bivariate normal selection model; MB-BC = bias-corrected estimator using $\theta = 0.0.5$ or 0.25; and, P* is the bias-minimizing propensity score.

Table 5. Effect of SBP Participation on the Probability of Being Overweight

	Third Grade Outcome			Fifth Grade Outcome		
	Specification (1)	Specification (2)	Specification (3)	Specification (1)	Specification (2)	Specification (3)
A. ATE						
τ_{OLS}	0.036	0.035	0.035	0.050	0.050	0.049
	[0.015, 0.053]	[0.013, 0.053]	[0.014, 0.054]	[0.035, 0.081]	[0.029, 0.078]	[0.032, 0.076]
τ_{HI}	0.061	0.048	0.048	0.081	0.081	0.076
	[0.036, 0.090]	[0.024, 0.077]	[0.027, 0.077]	[0.058, 0.117]	[0.045, 0.107]	[0.044, 0.106]
$\tau_{MB,0.05}$	0.034	0.054	0.003	0.058	0.037	0.057
	[-0.059, 0.088]	[-0.064, 0.083]	[-0.058, 0.094]	[-0.044, 0.119]	[-0.051, 0.122]	[-0.053, 0.116]
$\tau_{MB,0.25}$	0.025	0.011	0.016	0.034	0.035	0.030
	[-0.016, 0.038]	[-0.023, 0.042]	[-0.013, 0.049]	[-0.002, 0.065]	[-0.008, 0.069]	[-0.003, 0.065]
τ_{KV}	-0.144	-0.143	-0.136	-0.175	-0.197	-0.206
	[-0.282, 0.014]	[-0.274, 0.017]	[-0.283, 0.059]	[-0.320, 0.008]	[-0.353, 0.053]	[-0.359, 0.038]
τ_{BVN}	-0.231	-0.190	-0.208	-0.318	-0.279	-0.278
	[-0.388, -0.093]	[-0.333, -0.038]	[-0.329, -0.044]	[-0.460, -0.137]	[-0.418, -0.070]	[-0.397, -0.095]
$\tau_{MB\text{-}BC,0.05}$	-0.217	-0.159	-0.226	-0.295	-0.278	-0.255
	[-0.429, -0.033]	[-0.348, -0.023]	[-0.361, -0.019]	[-0.479, -0.094]	[-0.431, -0.035]	[-0.435, -0.060]
$\tau_{MB\text{-}BC,0.25}$	-0.226	-0.201	-0.213	-0.318	-0.281	-0.282
	[-0.396, -0.065]	[-0.331, -0.022]	[-0.332, -0.031]	[-0.478, -0.130]	[-0.424, -0.050]	[-0.397, -0.086]
P*	0.538	0.514	0.502	0.538	0.526	0.534
	[0.372, 0.874]	[0.314, 0.880]	[0.314, 0.835]	[0.392, 0.812]	[0.343, 0.823]	[0.365, 0.818]
B. ATT						
τ_{OLS}	0.036	0.035	0.035	0.050	0.050	0.049
	[0.015, 0.053]	[0.013, 0.053]	[0.014, 0.054]	[0.035, 0.081]	[0.029, 0.078]	[0.032, 0.076]
τ_{HI}	0.038	0.037	0.034	0.055	0.051	0.049
	[0.016, 0.067]	[0.012, 0.062]	[0.009, 0.061]	[0.029, 0.086]	[0.023, 0.081]	[0.025, 0.078]
$\tau_{MB,0.05}$	0.009	0.020	-0.002	0.030	0.022	0.008
	[-0.057, 0.083]	[-0.044, 0.086]	[-0.044, 0.097]	[-0.050, 0.105]	[-0.051, 0.109]	[-0.053, 0.110]
$\tau_{MB,0.25}$	0.017	0.020	0.018	0.035	0.040	0.041
	[-0.019, 0.040]	[-0.015, 0.045]	[-0.018, 0.043]	[0.013, 0.077]	[0.009, 0.081]	[0.007, 0.077]
τ_{KV}	-0.144	-0.143	-0.136	-0.175	-0.197	-0.206
	[-0.282, 0.014]	[-0.274, 0.017]	[-0.283, 0.059]	[-0.320, 0.008]	[-0.353, 0.053]	[-0.359, 0.038]
τ_{BVN}	-0.094	-0.077	-0.088	-0.122	-0.107	-0.108
	[-0.183, 0.000]	[-0.156, 0.018]	[-0.163, 0.007]	[-0.210, -0.021]	[-0.191, 0.010]	[-0.182, -0.004]
$\tau_{MB\text{-}BC,0.05}$	-0.234	-0.190	-0.230	-0.311	-0.286	-0.296
	[-0.414, -0.041]	[-0.352, 0.003]	[-0.357, -0.021]	[-0.491, -0.084]	[-0.459, -0.018]	[-0.453, -0.027]
$\tau_{MB\text{-}BC,0.25}$	-0.226	-0.190	-0.211	-0.306	-0.269	-0.263
	[-0.411, -0.052]	[-0.358, -0.009]	[-0.361, -0.021]	[-0.472, -0.090]	[-0.433, -0.021]	[-0.413, -0.053]
P*	0.500	0.500	0.500	0.500	0.500	0.500

Notes: Overweight is defined as BMI above the 85th percentile. For other details, see Table 4.

Table 6. Effect of SBP Participation on the Probability of Being Obese

	Third Grade Outcome			Fifth Grade Outcome		
	Specification (1)	Specification (2)	Specification (3)	Specification (1)	Specification (2)	Specification (3)
A. ATE						
τ_{OLS}	0.027	0.028	0.028	0.039	0.039	0.039
	[0.007, 0.038]	[0.006, 0.041]	[0.008, 0.043]	[0.019, 0.064]	[0.022, 0.066]	[0.024, 0.062]
τ_{HI}	0.034	0.029	0.028	0.054	0.059	0.053
	[0.011, 0.056]	[0.008, 0.052]	[0.008, 0.049]	[0.030, 0.084]	[0.028, 0.087]	[0.026, 0.080]
$\tau_{MB,0.05}$	0.002	0.026	0.001	0.050	0.062	0.068
	[-0.063, 0.076]	[-0.055, 0.068]	[-0.063, 0.080]	[-0.045, 0.108]	[-0.046, 0.110]	[-0.031, 0.103]
$\tau_{MB,0.25}$	0.011	0.021	0.015	0.047	0.048	0.041
	[-0.017, 0.035]	[-0.016, 0.039]	[-0.015, 0.043]	[-0.009, 0.064]	[-0.002, 0.065]	[-0.005, 0.063]
τ_{KV}	-0.185	-0.150	-0.147	-0.119	-0.076	-0.098
	[-0.300, -0.058]	[-0.276, -0.005]	[-0.266, 0.006]	[-0.265, 0.029]	[-0.203, 0.089]	[-0.245, 0.066]
τ_{BVN}	-0.210	-0.158	-0.166	-0.192	-0.118	-0.134
	[-0.314, -0.103]	[-0.275, -0.032]	[-0.251, -0.026]	[-0.313, -0.040]	[-0.253, 0.038]	[-0.276, 0.015]
$\tau_{MB\text{-}BC,0.05}$	-0.222	-0.148	-0.182	-0.170	-0.088	-0.096
	[-0.341, -0.077]	[-0.304, -0.011]	[-0.288, -0.005]	[-0.350, 0.004]	[-0.283, 0.059]	[-0.293, 0.061]
$\tau_{MB\text{-}BC,0.25}$	-0.213	-0.154	-0.167	-0.173	-0.102	-0.123
	[-0.316, -0.097]	[-0.287, -0.027]	[-0.257, -0.011]	[-0.321, -0.026]	[-0.260, 0.037]	[-0.279, 0.044]
P*	0.474	0.434	0.434	0.562	0.570	0.550
	[0.321, 0.728]	[0.243, 0.796]	[0.253, 0.850]	[0.361, 0.886]	[0.188, 0.943]	[0.199, 0.886]
B. ATT						
τ_{OLS}	0.027	0.028	0.028	0.039	0.039	0.039
	[0.007, 0.038]	[0.006, 0.041]	[0.008, 0.043]	[0.019, 0.064]	[0.022, 0.066]	[0.024, 0.062]
τ_{HI}	0.033	0.031	0.032	0.049	0.045	0.044
	[0.013, 0.051]	[0.007, 0.051]	[0.008, 0.052]	[0.021, 0.075]	[0.019, 0.071]	[0.022, 0.070]
$\tau_{MB,0.05}$	-0.003	0.012	0.016	0.059	0.025	0.044
	[-0.045, 0.071]	[-0.038, 0.081]	[-0.037, 0.105]	[-0.027, 0.112]	[-0.042, 0.117]	[-0.024, 0.119]
$\tau_{MB,0.25}$	0.020	0.021	0.023	0.039	0.041	0.043
	[-0.010, 0.039]	[-0.010, 0.038]	[-0.003, 0.048]	[0.006, 0.072]	[0.006, 0.071]	[0.011, 0.075]
τ_{KV}	-0.185	-0.150	-0.147	-0.119	-0.076	-0.098
	[-0.300, -0.058]	[-0.276, -0.005]	[-0.266, 0.006]	[-0.265, 0.029]	[-0.203, 0.089]	[-0.245, 0.066]
τ_{BVN}	-0.094	-0.071	-0.076	-0.068	-0.033	-0.043
	[-0.163, -0.034]	[-0.144, 0.000]	[-0.128, 0.009]	[-0.141, 0.016]	[-0.110, 0.059]	[-0.130, 0.047]
$\tau_{MB\text{-}BC,0.05}$	-0.232	-0.173	-0.178	-0.151	-0.117	-0.114
	[-0.365, -0.068]	[-0.310, 0.001]	[-0.277, 0.045]	[-0.339, -0.001]	[-0.284, 0.102]	[-0.301, 0.082]
$\tau_{MB\text{-}BC,0.25}$	-0.209	-0.164	-0.171	-0.171	-0.101	-0.115
	[-0.346, -0.089]	[-0.302, -0.025]	[-0.273, -0.006]	[-0.326, -0.003]	[-0.257, 0.091]	[-0.292, 0.070]
P*	0.500	0.500	0.500	0.500	0.500	0.500

Notes: Obese is defined as BMI above the 95th percentile. For other details, see Table 4.

Table 7. Effect of SBP Participation on the Probability of Being Underweight

	Third Grade Outcome			Fifth Grade Outcome		
	Specification (1)	Specification (2)	Specification (3)	Specification (1)	Specification (2)	Specification (3)
A. ATE						
τ_{OLS}	-0.020	-0.021	-0.020	-0.017	-0.018	-0.017
	[-0.031, -0.010]	[-0.031, -0.010]	[-0.029, -0.007]	[-0.029, -0.004]	[-0.030, -0.004]	[-0.028, -0.001]
τ_{HI}	-0.026	-0.025	-0.025	-0.032	-0.027	-0.025
	[-0.041, -0.015]	[-0.037, -0.012]	[-0.037, -0.010]	[-0.046, -0.015]	[-0.039, -0.011]	[-0.037, -0.008]
$\tau_{MB,0.05}$	-0.016	-0.018	-0.021	0.006	0.000	-0.004
	[-0.056, 0.021]	[-0.058, 0.024]	[-0.056, 0.022]	[-0.060, 0.035]	[-0.057, 0.028]	[-0.059, 0.030]
$\tau_{MB,0.25}$	-0.017	-0.016	-0.016	-0.010	-0.011	-0.011
	[-0.040, -0.000]	[-0.046, -0.001]	[-0.039, 0.001]	[-0.030, 0.006]	[-0.037, 0.005]	[-0.036, 0.010]
τ_{KV}	-0.007	-0.018	0.003	0.091	0.036	0.054
	[-0.085, 0.071]	[-0.105, 0.096]	[-0.102, 0.102]	[-0.005, 0.183]	[-0.073, 0.143]	[-0.065, 0.146]
τ_{BVN}	0.018	0.002	0.016	0.115	0.048	0.054
	[-0.072, 0.104]	[-0.091, 0.098]	[-0.072, 0.096]	[0.013, 0.212]	[-0.057, 0.149]	[-0.046, 0.140]
$\tau_{MB\text{-}BC,0.05}$	0.019	-0.003	0.010	0.129	0.054	0.055
	[-0.077, 0.107]	[-0.089, 0.111]	[-0.079, 0.110]	[0.009, 0.215]	[-0.085, 0.144]	[-0.057, 0.144]
$\tau_{MB\text{-}BC,0.25}$	0.017	0.000	0.015	0.113	0.042	0.048
	[-0.073, 0.109]	[-0.092, 0.098]	[-0.061, 0.094]	[-0.006, 0.209]	[-0.059, 0.145]	[-0.044, 0.139]
P*	0.679	0.880	0.711	0.577	0.793	0.747
	[0.046, 0.961]	[0.025, 0.961]	[0.023, 0.959]	[0.357, 0.926]	[0.145, 0.957]	[0.074, 0.972]
B. ATT						
τ_{OLS}	-0.020	-0.021	-0.020	-0.017	-0.018	-0.017
	[-0.031, -0.010]	[-0.031, -0.010]	[-0.029, -0.007]	[-0.029, -0.004]	[-0.030, -0.004]	[-0.028, -0.001]
τ_{HI}	-0.020	-0.017	-0.017	-0.020	-0.013	-0.011
	[-0.034, -0.007]	[-0.030, -0.006]	[-0.029, -0.005]	[-0.042, -0.000]	[-0.027, 0.002]	[-0.024, 0.004]
$\tau_{MB,0.05}$	-0.014	-0.024	-0.026	-0.019	-0.042	-0.042
	[-0.061, 0.015]	[-0.060, 0.020]	[-0.054, 0.020]	[-0.062, 0.023]	[-0.054, 0.029]	[-0.065, 0.020]
$\tau_{MB,0.25}$	-0.015	-0.018	-0.015	-0.009	-0.012	-0.015
	[-0.032, 0.001]	[-0.028, 0.002]	[-0.028, 0.004]	[-0.029, 0.009]	[-0.029, 0.007]	[-0.027, 0.008]
τ_{KV}	-0.007	-0.018	0.003	0.091	0.036	0.054
	[-0.085, 0.071]	[-0.105, 0.096]	[-0.102, 0.102]	[-0.005, 0.183]	[-0.073, 0.143]	[-0.065, 0.146]
τ_{BVN}	-0.003	-0.012	-0.004	0.044	0.008	0.012
	[-0.053, 0.052]	[-0.066, 0.046]	[-0.054, 0.044]	[-0.016, 0.101]	[-0.052, 0.069]	[-0.043, 0.066]
$\tau_{MB\text{-}BC,0.05}$	0.015	-0.013	0.000	0.097	0.000	0.007
	[-0.095, 0.114]	[-0.129, 0.111]	[-0.100, 0.109]	[-0.027, 0.219]	[-0.119, 0.137]	[-0.098, 0.133]
$\tau_{MB\text{-}BC,0.25}$	0.015	-0.007	0.012	0.106	0.029	0.034
	[-0.086, 0.115]	[-0.107, 0.111]	[-0.094, 0.107]	[-0.020, 0.226]	[-0.092, 0.151]	[-0.076, 0.138]
P*	0.500	0.500	0.500	0.500	0.500	0.500

Notes: Underweight is defined as BMI below the 20th percentile. For other details, see Table 4.

Table A1. Monte Carlo Results: Estimates in the Common Effect Model ($\rho_0\sigma_0 = 0$; $\rho_{01} = 1$)

	Homoskedastic Error in Treatment Equation								Heteroskedastic Error in Treatment Equation							
	ATE				ATT				ATE				ATT			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
I. Normally Distributed Errors																
τ_{HI}	0.617	0.047	0.047	0.048	0.564	0.053	0.052	0.053	0.619	0.047	0.043	0.043	0.570	0.048	0.044	0.045
$\tau_{MB,0.05}$	0.342	0.113	0.122	0.111	0.231	0.102	0.100	0.106	0.331	0.138	0.103	0.101	0.725	0.114	0.110	0.106
$\tau_{MB,0.25}$	0.383	0.064	0.055	0.056	0.587	0.056	0.054	0.053	0.357	0.060	0.056	0.057	0.771	0.060	0.052	0.053
$\tau_{MB,EE,0.05}$	0.281	0.129	0.129	0.129	0.380	0.123	0.134	0.124	0.571	0.176	0.121	0.118	0.361	0.169	0.115	0.113
$\tau_{MB,EE,0.25}$	0.537	0.068	0.067	0.068	0.389	0.067	0.071	0.071	0.671	0.071	0.061	0.062	0.458	0.070	0.061	0.062
τ_{KV}	3.312	0.360	0.428	0.406	3.312	0.360	0.428	0.406	3.238	0.238	0.316	0.309	3.238	0.238	0.316	0.309
τ_{BVN}	17.906	0.301	0.335	0.328	11.696	0.226	0.249	0.243	10.622	0.232	0.261	0.261	7.139	0.177	0.201	0.200
$\tau_{MB\text{-}BC,0.05}$	17.304	0.312	0.354	0.341	15.478	0.321	0.345	0.341	10.038	0.257	0.267	0.265	9.453	0.259	0.302	0.296
$\tau_{MB\text{-}BC,0.25}$	17.357	0.296	0.326	0.321	15.855	0.307	0.331	0.324	10.099	0.225	0.250	0.248	9.544	0.240	0.274	0.273
$\tau_{BVN,\ EE}$	195.750	1.610	1.543	1.514	102.382	1.068	1.078	1.052	533.206	0.828	0.863	0.823	321.761	0.548	0.749	0.711
$\tau_{MB\text{-}BC,EE,0.05}$	194.989	1.319	1.384	1.338	147.330	1.381	1.338	1.350	533.022	0.730	1.146	1.151	452.463	0.688	0.875	0.864
$\tau_{MB\text{-}BC,EE,0.25}$	194.718	1.289	1.352	1.309	147.343	1.348	1.315	1.323	533.118	0.698	1.118	1.131	452.561	0.673	0.852	0.843
II. Asymmetric, Non-Normally Distributed Errors																
τ_{HI}	0.615	0.050	0.047	0.048	0.562	0.055	0.052	0.051	0.616	0.048	0.043	0.043	0.567	0.051	0.047	0.047
$\tau_{MB,0.05}$	0.332	0.097	0.107	0.105	0.218	0.100	0.105	0.113	0.385	0.125	0.109	0.111	0.728	0.112	0.103	0.111
$\tau_{MB,0.25}$	0.377	0.059	0.056	0.055	0.587	0.055	0.052	0.055	0.386	0.064	0.056	0.057	0.763	0.059	0.054	0.051
$\tau_{MB,EE,0.05}$	0.291	0.130	0.127	0.128	0.357	0.128	0.127	0.126	0.596	0.171	0.128	0.133	0.344	0.161	0.125	0.123
$\tau_{MB,EE,0.25}$	0.548	0.069	0.067	0.068	0.400	0.069	0.069	0.070	0.691	0.069	0.059	0.064	0.455	0.067	0.063	0.063
τ_{KV}	3.308	0.363	0.441	0.408	3.308	0.363	0.441	0.408	3.233	0.248	0.329	0.318	3.233	0.248	0.329	0.318
τ_{BVN}	17.093	0.281	0.322	0.315	11.200	0.207	0.236	0.232	10.158	0.245	0.269	0.260	6.842	0.184	0.204	0.198
$\tau_{MB\text{-}BC,0.05}$	16.487	0.286	0.340	0.329	14.768	0.289	0.329	0.338	9.612	0.273	0.285	0.279	9.061	0.257	0.290	0.288
$\tau_{MB\text{-}BC,0.25}$	16.542	0.277	0.319	0.311	15.159	0.278	0.312	0.309	9.657	0.242	0.265	0.255	9.142	0.248	0.276	0.267
$\tau_{BVN,\ EE}$	171.703	1.608	1.610	1.526	89.061	1.071	1.105	1.047	501.755	0.899	0.923	0.866	303.361	0.598	0.786	0.736
$\tau_{MB\text{-}BC,EE,0.05}$	170.746	1.329	1.435	1.373	128.059	1.401	1.419	1.322	501.488	0.838	0.995	0.917	426.283	0.787	0.894	0.853
$\tau_{MB\text{-}BC,EE,0.25}$	170.477	1.296	1.405	1.342	128.073	1.377	1.390	1.285	501.588	0.815	0.970	0.894	426.399	0.762	0.878	0.833

Notes: Specification (1) is under-specified (relevant higher order terms are excluded); specification (2) is correctly specified; specification (3) is over-specified (irrelevant higher order terms are included); specification (4) is over-specified (irrelevant higher order terms and an irrelevant variable are included). Shading indicates best performance within each column. See text and Table 1 for further details.

Table A2. Monte Carlo Results: Estimates in the Common Effect Model ($\rho_0\sigma_0$ = -0.25; ρ_{01} = 1)

	Homoskedastic Error in Treatment Equation								Heteroskedastic Error in Treatment Equation							
	ATE				ATT				ATE				ATT			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
I. Normally Distributed Errors																
τ_{HI}	0.962	0.440	0.441	0.441	0.915	0.443	0.446	0.446	0.959	0.415	0.442	0.442	0.914	0.421	0.441	0.441
$\tau_{MB,0.05}$	0.724	0.407	0.399	0.403	0.582	0.411	0.413	0.409	0.742	0.406	0.410	0.405	1.058	0.432	0.411	0.411
$\tau_{MB,0.25}$	0.764	0.402	0.401	0.403	0.922	0.404	0.404	0.404	0.762	0.410	0.406	0.405	1.098	0.428	0.405	0.404
$\tau_{MB,EE,0.05}$	0.576	0.471	0.459	0.468	0.757	0.464	0.467	0.474	0.916	0.388	0.454	0.457	0.734	0.420	0.470	0.455
$\tau_{MB,EE,0.25}$	0.858	0.437	0.434	0.438	0.758	0.437	0.451	0.450	1.001	0.411	0.428	0.429	0.819	0.421	0.448	0.442
τ_{KV}	3.308	0.374	0.463	0.439	3.308	0.374	0.463	0.439	3.239	0.279	0.372	0.359	3.239	0.279	0.372	0.359
τ_{BVN}	15.800	0.287	0.324	0.322	10.481	0.245	0.267	0.270	9.046	0.265	0.271	0.274	6.243	0.259	0.242	0.247
$\tau_{MB\text{-}BC,0.05}$	15.283	0.301	0.336	0.342	13.692	0.296	0.323	0.314	8.580	0.266	0.295	0.291	8.132	0.294	0.289	0.291
$\tau_{MB\text{-}BC,0.25}$	15.325	0.291	0.326	0.322	14.037	0.283	0.313	0.310	8.619	0.264	0.277	0.277	8.209	0.278	0.273	0.276
$\tau_{BVN,\ EE}$	192.679	1.363	1.380	1.372	101.120	0.933	0.982	0.974	462.362	0.903	0.956	0.897	279.601	0.617	0.829	0.776
$\tau_{MB\text{-}BC,EE,0.05}$	192.262	1.268	1.435	1.494	145.840	1.185	1.219	1.205	462.119	0.823	1.213	1.196	392.971	0.761	0.965	0.963
$\tau_{MB\text{-}BC,EE,0.25}$	191.988	1.244	1.414	1.474	145.854	1.171	1.193	1.183	462.193	0.803	1.194	1.181	393.048	0.746	0.947	0.944
II. Asymmetric, Non-Normally Distributed Errors																
τ_{HI}	0.918	0.378	0.377	0.377	0.866	0.348	0.350	0.349	0.918	0.364	0.388	0.388	0.867	0.339	0.356	0.357
$\tau_{MB,0.05}$	0.628	0.350	0.354	0.352	0.619	0.420	0.419	0.420	0.847	0.338	0.369	0.368	1.074	0.439	0.418	0.425
$\tau_{MB,0.25}$	0.697	0.348	0.350	0.350	0.917	0.387	0.387	0.388	0.849	0.346	0.357	0.355	1.101	0.409	0.385	0.385
$\tau_{MB,EE,0.05}$	0.497	0.434	0.440	0.429	0.655	0.348	0.361	0.350	1.037	0.344	0.426	0.436	0.637	0.325	0.344	0.362
$\tau_{MB,EE,0.25}$	0.743	0.385	0.386	0.383	0.676	0.338	0.344	0.341	1.101	0.354	0.385	0.387	0.749	0.329	0.343	0.347
τ_{KV}	3.312	0.400	0.441	0.425	3.312	0.400	0.441	0.425	3.243	0.275	0.329	0.328	3.243	0.275	0.329	0.328
τ_{BVN}	14.677	0.308	0.348	0.341	9.828	0.300	0.317	0.314	8.226	0.288	0.292	0.293	5.748	0.300	0.305	0.307
$\tau_{MB\text{-}BC,0.05}$	14.108	0.313	0.343	0.344	12.891	0.381	0.407	0.404	7.838	0.294	0.310	0.301	7.544	0.395	0.389	0.396
$\tau_{MB\text{-}BC,0.25}$	14.179	0.303	0.338	0.333	13.195	0.350	0.376	0.372	7.874	0.285	0.289	0.288	7.606	0.363	0.358	0.361
$\tau_{BVN,\ EE}$	210.248	1.631	1.588	1.560	111.527	1.083	1.081	1.067	455.459	0.900	0.765	0.758	275.333	0.590	0.659	0.655
$\tau_{MB\text{-}BC,EE,0.05}$	210.027	1.514	1.592	1.447	160.829	1.389	1.370	1.326	454.976	0.903	1.177	1.130	387.099	0.774	0.762	0.751
$\tau_{MB\text{-}BC,EE,0.25}$	209.796	1.503	1.584	1.436	160.827	1.374	1.358	1.317	455.038	0.872	1.178	1.133	387.204	0.749	0.759	0.736

Notes: See Table A1.

Table A3. Monte Carlo Results: Estimates in the Common Effect Model ($\rho_0\sigma_0$ = -0.50; ρ_{01} = 1)

	Homoskedastic Error in Treatment Equation								Heteroskedastic Error in Treatment Equation							
	ATE				ATT				ATE				ATT			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
I. Normally Distributed Errors																
τ_{HI}	1.306	0.883	0.881	0.882	1.266	0.893	0.891	0.892	1.298	0.841	0.873	0.872	1.256	0.844	0.874	0.874
$\tau_{MB,0.05}$	1.109	0.795	0.796	0.803	0.967	0.800	0.801	0.798	1.130	0.817	0.798	0.805	1.388	0.817	0.800	0.808
$\tau_{MB,0.25}$	1.149	0.805	0.805	0.804	1.261	0.809	0.808	0.808	1.149	0.823	0.808	0.810	1.435	0.831	0.810	0.812
$\tau_{MB,EE,0.05}$	0.906	0.892	0.885	0.900	1.165	0.898	0.903	0.907	1.212	0.788	0.874	0.869	1.110	0.819	0.870	0.880
$\tau_{MB,EE,0.25}$	1.169	0.860	0.856	0.863	1.138	0.864	0.872	0.875	1.288	0.829	0.851	0.847	1.170	0.836	0.860	0.859
τ_{KV}	3.301	0.368	0.418	0.422	3.301	0.368	0.418	0.422	3.232	0.269	0.335	0.340	3.232	0.269	0.335	0.340
τ_{BVN}	14.099	0.278	0.299	0.302	9.518	0.312	0.328	0.338	8.264	0.300	0.271	0.276	5.853	0.394	0.335	0.345
$\tau_{MB\text{-}BC,0.05}$	13.665	0.287	0.322	0.324	12.280	0.294	0.308	0.307	7.874	0.317	0.286	0.303	7.493	0.334	0.290	0.294
$\tau_{MB\text{-}BC,0.25}$	13.703	0.285	0.308	0.312	12.574	0.277	0.291	0.294	7.913	0.314	0.280	0.286	7.567	0.328	0.276	0.282
$\tau_{BVN,\ EE}$	207.023	1.402	1.417	1.385	109.571	0.964	1.002	0.999	400.614	0.835	0.819	0.776	242.912	0.585	0.738	0.712
$\tau_{MB\text{-}BC,EE,0.05}$	206.831	1.422	1.526	1.482	158.241	1.247	1.248	1.220	400.363	0.791	1.045	1.065	341.154	0.731	0.922	0.853
$\tau_{MB\text{-}BC,EE,0.25}$	206.585	1.396	1.501	1.451	158.267	1.223	1.222	1.187	400.429	0.781	1.037	1.062	341.202	0.733	0.909	0.842
II. Asymmetric, Non-Normally Distributed Errors																
τ_{HI}	1.143	0.628	0.628	0.626	1.082	0.489	0.488	0.486	1.148	0.641	0.669	0.669	1.087	0.511	0.533	0.532
$\tau_{MB,0.05}$	1.288	0.391	0.361	0.357	1.087	0.784	0.781	0.781	1.411	0.348	0.475	0.475	1.411	0.806	0.783	0.788
$\tau_{MB,0.25}$	1.265	0.449	0.428	0.434	1.228	0.715	0.714	0.711	1.346	0.446	0.498	0.498	1.410	0.731	0.710	0.709
$\tau_{MB,EE,0.05}$	0.707	0.792	0.762	0.765	0.727	0.518	0.511	0.501	1.655	0.616	0.589	0.654	0.804	0.487	0.478	0.536
$\tau_{MB,EE,0.25}$	0.926	0.648	0.631	0.634	0.768	0.496	0.483	0.490	1.503	0.571	0.560	0.594	0.963	0.492	0.482	0.511
τ_{KV}	3.310	0.368	0.484	0.446	3.310	0.368	0.484	0.446	3.236	0.257	0.335	0.323	3.236	0.257	0.335	0.323
τ_{BVN}	7.641	0.380	0.370	0.372	5.635	0.585	0.528	0.533	3.816	0.364	0.424	0.426	3.036	0.572	0.647	0.648
$\tau_{MB\text{-}BC,0.05}$	7.333	0.378	0.389	0.383	7.321	0.868	0.812	0.820	3.775	0.373	0.460	0.457	3.968	0.866	0.966	0.972
$\tau_{MB\text{-}BC,0.25}$	7.373	0.396	0.376	0.380	7.465	0.801	0.748	0.753	3.732	0.390	0.467	0.466	4.004	0.794	0.896	0.896
$\tau_{BVN,\ EE}$	266.352	1.819	1.707	1.674	144.190	1.172	1.101	1.101	345.781	1.080	0.730	0.740	210.121	0.700	0.673	0.704
$\tau_{MB\text{-}BC,EE,0.05}$	266.189	1.764	1.810	1.599	208.051	1.514	1.418	1.403	343.166	1.169	0.872	0.902	295.415	0.956	0.755	0.819
$\tau_{MB\text{-}BC,EE,0.25}$	265.986	1.786	1.839	1.624	207.986	1.443	1.360	1.369	343.034	1.125	0.892	0.918	295.549	0.874	0.748	0.811

Notes: See Table A1.

Table A4. Monte Carlo Results: Estimates in the Heterogeneous Effect Model ($\rho_0\sigma_0 = 0$; $\rho_{01} = 0.50$; $\rho_\delta\sigma_\delta = 0$)

	Homoskedastic Error in Treatment Equation								Heteroskedastic Error in Treatment Equation							
	ATE				ATT				ATE				ATT			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
I. Normally Distributed Errors																
τ_{HI}	0.615	0.046	0.043	0.044	0.561	0.052	0.049	0.050	0.618	0.046	0.040	0.040	0.569	0.046	0.043	0.043
$\tau_{MB,0.05}$	0.345	0.100	0.107	0.103	0.232	0.108	0.107	0.110	0.356	0.125	0.103	0.116	0.731	0.115	0.112	0.110
$\tau_{MB,0.25}$	0.382	0.062	0.060	0.063	0.590	0.056	0.058	0.057	0.374	0.062	0.055	0.056	0.768	0.061	0.055	0.054
$\tau_{MB,EE,0.05}$	0.285	0.124	0.130	0.129	0.349	0.111	0.121	0.120	0.572	0.158	0.116	0.115	0.365	0.169	0.119	0.116
$\tau_{MB,EE,0.25}$	0.540	0.066	0.062	0.064	0.384	0.065	0.064	0.065	0.686	0.068	0.062	0.061	0.454	0.071	0.058	0.056
τ_{KV}	3.304	0.361	0.450	0.426	3.304	0.361	0.450	0.426	3.237	0.258	0.342	0.328	3.237	0.258	0.342	0.328
τ_{BVN}	17.500	0.288	0.325	0.322	11.449	0.213	0.239	0.237	10.454	0.253	0.271	0.268	7.034	0.191	0.208	0.206
$\tau_{MB\text{-}BC,0.05}$	16.904	0.299	0.339	0.335	15.157	0.298	0.324	0.326	9.898	0.276	0.275	0.280	9.350	0.283	0.299	0.295
$\tau_{MB\text{-}BC,0.25}$	16.951	0.284	0.319	0.313	15.538	0.284	0.318	0.315	9.945	0.246	0.260	0.259	9.415	0.252	0.278	0.274
$\tau_{BVN,\ EE}$	188.928	1.585	1.575	1.539	98.698	1.053	1.080	1.066	528.207	0.867	0.959	0.912	318.666	0.582	0.822	0.783
$\tau_{MB\text{-}BC,EE,0.05}$	187.976	1.389	1.413	1.442	142.308	1.383	1.360	1.357	528.023	0.794	1.086	1.067	448.563	0.731	0.945	0.926
$\tau_{MB\text{-}BC,EE,0.25}$	187.702	1.360	1.377	1.407	142.308	1.357	1.334	1.323	528.138	0.767	1.073	1.054	448.653	0.708	0.925	0.904
II. Asymmetric, Non-Normally Distributed Errors																
τ_{HI}	0.616	0.045	0.043	0.043	0.562	0.051	0.050	0.050	0.619	0.044	0.040	0.040	0.569	0.046	0.042	0.043
$\tau_{MB,0.05}$	0.339	0.111	0.107	0.105	0.221	0.102	0.108	0.113	0.367	0.127	0.107	0.103	0.729	0.106	0.105	0.110
$\tau_{MB,0.25}$	0.383	0.056	0.056	0.056	0.589	0.051	0.052	0.050	0.371	0.063	0.056	0.056	0.771	0.060	0.053	0.053
$\tau_{MB,EE,0.05}$	0.280	0.124	0.120	0.119	0.377	0.120	0.121	0.119	0.562	0.159	0.116	0.113	0.358	0.151	0.122	0.117
$\tau_{MB,EE,0.25}$	0.533	0.067	0.063	0.061	0.393	0.066	0.066	0.068	0.670	0.065	0.060	0.059	0.460	0.068	0.061	0.063
τ_{KV}	3.306	0.400	0.493	0.440	3.306	0.400	0.493	0.440	3.233	0.273	0.363	0.351	3.233	0.273	0.363	0.351
τ_{BVN}	17.484	0.298	0.328	0.318	11.438	0.221	0.242	0.234	10.559	0.254	0.282	0.275	7.105	0.192	0.215	0.209
$\tau_{MB\text{-}BC,0.05}$	16.882	0.310	0.342	0.321	15.115	0.305	0.334	0.327	10.003	0.281	0.288	0.281	9.418	0.280	0.307	0.308
$\tau_{MB\text{-}BC,0.25}$	16.934	0.291	0.320	0.310	15.503	0.295	0.320	0.309	10.041	0.250	0.270	0.265	9.499	0.260	0.287	0.280
$\tau_{BVN,\ EE}$	174.035	1.405	1.414	1.358	89.977	0.931	0.974	0.934	529.111	0.861	0.850	0.796	318.996	0.566	0.747	0.695
$\tau_{MB\text{-}BC,EE,0.05}$	173.154	1.235	1.312	1.312	129.674	1.248	1.223	1.185	528.875	0.831	1.130	1.017	449.411	0.754	0.878	0.793
$\tau_{MB\text{-}BC,EE,0.25}$	172.881	1.206	1.289	1.290	129.696	1.221	1.199	1.164	528.977	0.810	1.112	0.998	449.510	0.733	0.855	0.770

Notes: See Table A1.

Table A5. Monte Carlo Results: Estimates in the Heterogeneous Effect Model ($\rho_0\sigma_0$ = -0.20; ρ_{01} = 0.50; $\rho_\delta\sigma_\delta$ = -0.10)

	Homoskedastic Error in Treatment Equation								Heteroskedastic Error in Treatment Equation							
	ATE				ATT				ATE				ATT			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
I. Normally Distributed Errors																
τ_{HI}	0.935	0.412	0.412	0.413	0.842	0.358	0.357	0.358	0.933	0.390	0.415	0.416	0.843	0.338	0.358	0.359
$\tau_{MB,0.05}$	0.657	0.379	0.383	0.388	0.520	0.326	0.325	0.326	0.772	0.385	0.392	0.388	0.979	0.352	0.322	0.325
$\tau_{MB,0.25}$	0.720	0.378	0.375	0.376	0.828	0.308	0.307	0.306	0.785	0.378	0.377	0.374	1.007	0.327	0.303	0.302
$\tau_{MB,EE,0.05}$	0.534	0.463	0.468	0.466	0.619	0.333	0.342	0.340	0.938	0.368	0.448	0.447	0.593	0.293	0.338	0.338
$\tau_{MB,EE,0.25}$	0.790	0.413	0.410	0.412	0.618	0.304	0.312	0.314	1.036	0.380	0.407	0.409	0.694	0.290	0.313	0.314
τ_{KV}	3.316	0.369	0.461	0.440	3.272	0.370	0.466	0.440	3.231	0.254	0.321	0.315	3.189	0.262	0.332	0.322
τ_{BVN}	14.691	0.262	0.311	0.311	9.768	0.220	0.244	0.248	8.343	0.266	0.259	0.258	5.776	0.238	0.218	0.220
$\tau_{MB\text{-}BC,0.05}$	14.144	0.271	0.318	0.316	12.750	0.274	0.320	0.315	7.911	0.262	0.277	0.283	7.524	0.297	0.279	0.281
$\tau_{MB\text{-}BC,0.25}$	14.207	0.259	0.306	0.307	13.063	0.263	0.302	0.302	7.955	0.253	0.263	0.261	7.586	0.275	0.269	0.269
$\tau_{BVN,\ EE}$	192.651	1.708	1.621	1.558	101.254	1.145	1.121	1.072	450.245	0.868	0.813	0.769	272.382	0.575	0.709	0.671
$\tau_{MB\text{-}BC,EE,0.05}$	192.470	1.431	1.517	1.490	146.116	1.471	1.420	1.334	449.932	0.942	1.208	1.128	382.924	0.743	0.803	0.770
$\tau_{MB\text{-}BC,EE,0.25}$	192.225	1.401	1.499	1.474	146.130	1.450	1.391	1.315	450.023	0.916	1.200	1.125	383.016	0.737	0.785	0.754
II. Asymmetric, Non-Normally Distributed Errors																
τ_{HI}	0.908	0.380	0.379	0.379	0.806	0.287	0.287	0.287	0.907	0.366	0.390	0.391	0.821	0.287	0.309	0.309
$\tau_{MB,0.05}$	0.597	0.304	0.312	0.311	0.546	0.314	0.315	0.317	0.958	0.269	0.338	0.337	1.040	0.399	0.384	0.381
$\tau_{MB,0.25}$	0.702	0.311	0.313	0.315	0.820	0.272	0.272	0.271	0.944	0.305	0.330	0.329	1.065	0.362	0.340	0.338
$\tau_{MB,EE,0.05}$	0.472	0.496	0.482	0.492	0.539	0.242	0.234	0.231	1.140	0.417	0.478	0.478	0.579	0.294	0.303	0.314
$\tau_{MB,EE,0.25}$	0.709	0.397	0.392	0.395	0.544	0.210	0.212	0.209	1.169	0.368	0.391	0.396	0.697	0.275	0.284	0.285
τ_{KV}	3.286	0.380	0.446	0.422	3.241	0.391	0.453	0.426	3.209	0.270	0.337	0.315	3.178	0.279	0.345	0.321
τ_{BVN}	12.827	0.279	0.342	0.340	8.675	0.215	0.244	0.247	7.240	0.252	0.253	0.250	5.129	0.245	0.248	0.248
$\tau_{MB\text{-}BC,0.05}$	12.233	0.291	0.342	0.341	11.311	0.281	0.322	0.329	6.962	0.285	0.267	0.262	6.749	0.342	0.350	0.351
$\tau_{MB\text{-}BC,0.25}$	12.336	0.280	0.336	0.332	11.593	0.256	0.296	0.301	6.980	0.254	0.260	0.258	6.809	0.317	0.310	0.310
$\tau_{BVN,\ EE}$	221.506	1.641	1.592	1.551	117.753	1.075	1.063	1.037	422.570	0.953	0.819	0.793	256.358	0.630	0.708	0.687
$\tau_{MB\text{-}BC,EE,0.05}$	221.399	1.472	1.452	1.426	169.958	1.377	1.350	1.280	421.728	0.992	1.168	1.059	360.191	0.790	0.819	0.775
$\tau_{MB\text{-}BC,EE,0.25}$	221.175	1.484	1.453	1.431	169.955	1.369	1.332	1.263	421.760	0.976	1.182	1.069	360.294	0.771	0.816	0.774

Notes: See Table A1.

Table A6. Monte Carlo Results: Estimates in the Heterogeneous Effect Model ($\rho_0\sigma_0$ = -0.40; ρ_{01} = 0.50; $\rho_\delta\sigma_\delta$ = -0.10)

	Homoskedastic Error in Treatment Equation								Heteroskedastic Error in Treatment Equation							
	ATE				ATT				ATE				ATT			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
I. Normally Distributed Errors																
τ_{HI}	1.217	0.768	0.769	0.769	1.128	0.719	0.720	0.719	1.209	0.734	0.768	0.767	1.122	0.683	0.715	0.715
$\tau_{MB,0.05}$	0.962	0.713	0.718	0.716	0.830	0.653	0.659	0.654	1.146	0.730	0.701	0.704	1.262	0.680	0.656	0.659
$\tau_{MB,0.25}$	1.029	0.709	0.710	0.711	1.116	0.645	0.644	0.644	1.135	0.725	0.705	0.705	1.290	0.662	0.641	0.641
$\tau_{MB,EE,0.05}$	0.826	0.810	0.814	0.815	0.950	0.697	0.703	0.708	1.231	0.693	0.776	0.781	0.923	0.640	0.697	0.698
$\tau_{MB,EE,0.25}$	1.061	0.756	0.760	0.759	0.941	0.671	0.678	0.674	1.296	0.719	0.747	0.745	0.999	0.644	0.671	0.668
τ_{KV}	3.299	0.369	0.455	0.442	3.256	0.373	0.452	0.434	3.228	0.268	0.359	0.345	3.187	0.275	0.363	0.344
τ_{BVN}	13.197	0.261	0.301	0.305	8.918	0.266	0.301	0.311	7.527	0.284	0.267	0.270	5.343	0.325	0.286	0.294
$\tau_{MB\text{-}BC,0.05}$	12.705	0.264	0.321	0.326	11.520	0.271	0.308	0.312	7.208	0.302	0.284	0.292	6.875	0.321	0.285	0.299
$\tau_{MB\text{-}BC,0.25}$	12.772	0.264	0.307	0.309	11.808	0.263	0.295	0.302	7.223	0.289	0.270	0.274	6.929	0.299	0.272	0.273
$\tau_{BVN,\ EE}$	211.512	1.434	1.514	1.424	112.134	0.977	1.054	1.007	403.337	0.852	0.832	0.786	244.920	0.586	0.731	0.701
$\tau_{MB\text{-}BC,EE,0.05}$	211.378	1.375	1.471	1.343	162.320	1.243	1.321	1.274	402.984	0.973	1.069	1.003	343.887	0.802	0.888	0.833
$\tau_{MB\text{-}BC,EE,0.25}$	211.160	1.362	1.448	1.324	162.326	1.214	1.295	1.249	403.036	0.962	1.063	1.000	343.953	0.789	0.874	0.826
II. Asymmetric, Non-Normally Distributed Errors																
τ_{HI}	1.095	0.604	0.604	0.604	0.997	0.443	0.444	0.444	1.099	0.607	0.635	0.634	1.014	0.465	0.486	0.486
$\tau_{MB,0.05}$	1.202	0.325	0.322	0.327	1.004	0.695	0.692	0.689	1.340	0.238	0.395	0.404	1.345	0.727	0.706	0.698
$\tau_{MB,0.25}$	1.195	0.401	0.401	0.402	1.156	0.632	0.630	0.628	1.283	0.400	0.445	0.448	1.347	0.662	0.640	0.639
$\tau_{MB,EE,0.05}$	0.660	0.774	0.791	0.794	0.714	0.443	0.475	0.461	1.603	0.609	0.660	0.687	0.768	0.470	0.493	0.479
$\tau_{MB,EE,0.25}$	0.878	0.621	0.629	0.631	0.732	0.440	0.446	0.441	1.441	0.554	0.587	0.597	0.923	0.465	0.471	0.464
τ_{KV}	3.252	0.437	0.491	0.442	3.214	0.454	0.504	0.449	3.174	0.316	0.398	0.371	3.149	0.328	0.408	0.378
τ_{BVN}	7.792	0.289	0.372	0.357	5.668	0.322	0.299	0.307	3.597	0.250	0.287	0.289	2.851	0.382	0.432	0.438
$\tau_{MB\text{-}BC,0.05}$	7.460	0.329	0.406	0.393	7.403	0.520	0.496	0.506	3.589	0.303	0.318	0.327	3.750	0.623	0.684	0.685
$\tau_{MB\text{-}BC,0.25}$	7.521	0.300	0.369	0.355	7.555	0.464	0.441	0.451	3.530	0.278	0.339	0.345	3.784	0.561	0.621	0.627
$\tau_{BVN,\ EE}$	271.652	1.772	1.731	1.675	146.545	1.155	1.135	1.107	347.390	0.915	0.757	0.744	211.714	0.631	0.677	0.679
$\tau_{MB\text{-}BC,EE,0.05}$	271.501	1.797	1.838	1.629	211.451	1.486	1.436	1.396	344.602	1.031	1.174	1.196	297.136	0.827	0.797	0.788
$\tau_{MB\text{-}BC,EE,0.25}$	271.297	1.860	1.897	1.675	211.408	1.456	1.399	1.373	344.461	1.014	1.216	1.252	297.267	0.779	0.776	0.772

Notes: See Table A1.

Table B1. Monte Carlo Results: Estimates in the Common Effect Model ($\rho_0\sigma_0 = 0$; $\rho_{01} = 1$)

	Homoskedastic Error in Treatment Equation								Heteroskedastic Error in Treatment Equation							
	ATE				ATT				ATE				ATT			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
I. Normally Distributed Errors																
τ_{HI}	0.614	0.006	0.007	0.006	0.560	0.008	0.008	0.008	0.616	0.014	0.011	0.011	0.566	0.009	0.009	0.009
$\tau_{MB,0.05}$	0.312	0.017	0.016	0.016	0.199	0.019	0.020	0.018	0.222	0.050	0.022	0.023	0.722	0.032	0.018	0.018
$\tau_{MB,0.25}$	0.380	0.008	0.008	0.008	0.591	0.007	0.007	0.007	0.284	0.023	0.013	0.013	0.770	0.027	0.007	0.007
$\tau_{MB,EE,0.05}$	0.278	0.020	0.018	0.018	0.416	0.016	0.019	0.018	0.389	0.093	0.030	0.032	0.341	0.098	0.026	0.025
$\tau_{MB,EE,0.25}$	0.532	0.011	0.009	0.009	0.329	0.009	0.009	0.009	0.536	0.019	0.015	0.016	0.429	0.030	0.012	0.012
τ_{KV}	3.310	0.051	0.053	0.053	3.310	0.051	0.053	0.053	3.237	0.033	0.043	0.043	3.237	0.033	0.043	0.043
τ_{BVN}	17.237	0.037	0.040	0.039	11.295	0.028	0.030	0.029	10.181	0.036	0.039	0.039	6.878	0.028	0.031	0.031
$\tau_{MB\text{-}BC,0.05}$	16.616	0.040	0.042	0.043	14.915	0.043	0.047	0.046	9.563	0.061	0.045	0.045	9.163	0.049	0.049	0.048
$\tau_{MB\text{-}BC,0.25}$	16.683	0.037	0.039	0.039	15.308	0.037	0.040	0.040	9.625	0.042	0.038	0.038	9.211	0.043	0.042	0.041
$\tau_{BVN,\ EE}$	164.133	0.257	0.253	0.252	83.643	0.170	0.168	0.167	475.238	0.122	0.127	0.127	288.444	0.079	0.112	0.112
$\tau_{MB\text{-}BC,EE,0.05}$	163.422	0.204	0.217	0.228	120.067	0.216	0.211	0.211	475.021	0.134	0.177	0.180	404.771	0.145	0.130	0.132
$\tau_{MB\text{-}BC,EE,0.25}$	163.145	0.199	0.213	0.223	120.154	0.214	0.211	0.210	475.168	0.091	0.174	0.176	404.858	0.104	0.125	0.128
II. Asymmetric, Non-Normally Distributed Errors																
τ_{HI}	0.613	0.006	0.006	0.006	0.559	0.007	0.007	0.007	0.615	0.016	0.008	0.008	0.565	0.010	0.007	0.007
$\tau_{MB,0.05}$	0.312	0.014	0.016	0.015	0.199	0.015	0.015	0.015	0.221	0.063	0.020	0.021	0.732	0.032	0.014	0.013
$\tau_{MB,0.25}$	0.379	0.007	0.008	0.008	0.592	0.007	0.007	0.007	0.281	0.024	0.012	0.012	0.770	0.025	0.007	0.007
$\tau_{MB,EE,0.05}$	0.299	0.017	0.016	0.018	0.414	0.018	0.017	0.017	0.389	0.105	0.026	0.026	0.342	0.092	0.028	0.027
$\tau_{MB,EE,0.25}$	0.551	0.009	0.008	0.008	0.327	0.009	0.009	0.009	0.537	0.022	0.013	0.013	0.431	0.027	0.013	0.013
τ_{KV}	3.307	0.050	0.058	0.058	3.307	0.050	0.058	0.058	3.236	0.039	0.047	0.047	3.236	0.039	0.047	0.047
τ_{BVN}	17.324	0.034	0.040	0.040	11.346	0.024	0.029	0.028	10.064	0.033	0.035	0.035	6.801	0.024	0.026	0.026
$\tau_{MB\text{-}BC,0.05}$	16.704	0.037	0.043	0.043	14.991	0.033	0.039	0.038	9.449	0.068	0.038	0.038	9.074	0.042	0.036	0.035
$\tau_{MB\text{-}BC,0.25}$	16.771	0.032	0.039	0.039	15.384	0.032	0.038	0.038	9.510	0.037	0.034	0.035	9.113	0.039	0.035	0.035
$\tau_{BVN,\ EE}$	165.037	0.215	0.214	0.215	84.101	0.140	0.143	0.144	486.221	0.125	0.125	0.123	294.818	0.083	0.109	0.108
$\tau_{MB\text{-}BC,EE,0.05}$	164.240	0.177	0.187	0.188	120.755	0.179	0.182	0.183	485.999	0.155	0.132	0.128	413.877	0.139	0.134	0.147
$\tau_{MB\text{-}BC,EE,0.25}$	163.967	0.172	0.184	0.185	120.842	0.176	0.177	0.178	486.146	0.120	0.128	0.124	413.966	0.117	0.129	0.142

Notes: Figures represent root mean squared errors. Results based on 50 data sets with 250,000 observations each. Specification (1) is under-specified (relevant higher order terms are excluded); specifica (2) is correctly specified; specification (3) is over-specified (irrelevant higher order terms are included); specification (4) is over-specified (irrelevant higher order terms and an irrelevant variable are included). Shading indicates best performance within each column. See text and Table 1 for further details.

Table B2. Monte Carlo Results: Estimates in the Common Effect Model ($\rho_0\sigma_0$ = -0.25; ρ_{01} = 1)

	Homoskedastic Error in Treatment Equation								Heteroskedastic Error in Treatment Equation							
	ATE				ATT				ATE				ATT			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
I. Normally Distributed Errors																
τ_{HI}	0.960	0.442	0.441	0.441	0.912	0.446	0.446	0.446	0.957	0.416	0.442	0.442	0.911	0.419	0.441	0.441
$\tau_{MB,0.05}$	0.707	0.399	0.398	0.398	0.577	0.398	0.399	0.399	0.630	0.426	0.401	0.400	1.072	0.425	0.401	0.400
$\tau_{MB,0.25}$	0.765	0.404	0.404	0.404	0.920	0.405	0.405	0.405	0.679	0.427	0.403	0.403	1.105	0.428	0.405	0.405
$\tau_{MB,EE,0.05}$	0.508	0.443	0.436	0.438	0.807	0.428	0.427	0.426	0.769	0.384	0.433	0.432	0.735	0.413	0.422	0.419
$\tau_{MB,EE,0.25}$	0.768	0.424	0.421	0.422	0.722	0.418	0.418	0.418	0.876	0.416	0.420	0.419	0.800	0.425	0.418	0.417
τ_{KV}	3.310	0.048	0.054	0.054	3.310	0.048	0.054	0.054	3.237	0.039	0.043	0.043	3.237	0.039	0.043	0.043
τ_{BVN}	15.671	0.036	0.040	0.040	10.410	0.116	0.120	0.120	8.983	0.089	0.049	0.050	6.225	0.168	0.131	0.132
$\tau_{MB\text{-}BC,0.05}$	15.136	0.040	0.043	0.042	13.604	0.037	0.040	0.039	8.463	0.113	0.051	0.051	8.176	0.118	0.053	0.053
$\tau_{MB\text{-}BC,0.25}$	15.194	0.038	0.043	0.043	13.948	0.036	0.040	0.040	8.512	0.114	0.052	0.052	8.210	0.120	0.053	0.053
$\tau_{BVN,\ EE}$	179.771	0.204	0.199	0.198	93.074	0.166	0.163	0.164	427.639	0.121	0.121	0.121	260.080	0.131	0.149	0.149
$\tau_{MB\text{-}BC,EE,0.05}$	179.642	0.306	0.309	0.302	134.008	0.199	0.197	0.199	427.429	0.165	0.131	0.124	364.700	0.153	0.176	0.177
$\tau_{MB\text{-}BC,EE,0.25}$	179.383	0.313	0.314	0.306	134.092	0.198	0.194	0.196	427.535	0.125	0.133	0.124	364.765	0.126	0.172	0.174
II. Asymmetric, Non-Normally Distributed Errors																
τ_{HI}	0.920	0.379	0.379	0.379	0.868	0.346	0.346	0.346	0.920	0.365	0.391	0.391	0.870	0.336	0.355	0.355
$\tau_{MB,0.05}$	0.590	0.312	0.309	0.309	0.612	0.400	0.399	0.398	0.687	0.160	0.327	0.327	1.071	0.422	0.401	0.401
$\tau_{MB,0.25}$	0.697	0.323	0.324	0.324	0.926	0.383	0.383	0.383	0.735	0.294	0.337	0.337	1.102	0.405	0.382	0.382
$\tau_{MB,EE,0.05}$	0.474	0.318	0.316	0.314	0.674	0.315	0.333	0.337	1.034	0.128	0.325	0.325	0.652	0.242	0.350	0.334
$\tau_{MB,EE,0.25}$	0.716	0.332	0.332	0.332	0.623	0.326	0.333	0.335	1.175	0.294	0.337	0.337	0.742	0.302	0.345	0.339
τ_{KV}	3.309	0.054	0.056	0.056	3.309	0.054	0.056	0.056	3.237	0.035	0.044	0.044	3.237	0.035	0.044	0.044
τ_{BVN}	14.125	0.088	0.062	0.062	9.495	0.206	0.188	0.188	7.849	0.121	0.119	0.119	5.546	0.228	0.230	0.230
$\tau_{MB\text{-}BC,0.05}$	13.518	0.073	0.053	0.053	12.432	0.229	0.213	0.212	7.421	0.059	0.131	0.131	7.317	0.283	0.269	0.269
$\tau_{MB\text{-}BC,0.25}$	13.625	0.085	0.067	0.067	12.745	0.212	0.196	0.196	7.470	0.095	0.142	0.142	7.348	0.266	0.251	0.250
$\tau_{BVN,\ EE}$	194.729	0.223	0.220	0.220	101.804	0.185	0.187	0.187	407.323	0.131	0.110	0.108	248.174	0.127	0.157	0.156
$\tau_{MB\text{-}BC,EE,0.05}$	194.579	0.236	0.237	0.227	146.511	0.211	0.203	0.212	406.960	0.227	0.158	0.155	347.828	0.169	0.158	0.164
$\tau_{MB\text{-}BC,EE,0.25}$	194.338	0.243	0.249	0.232	146.561	0.219	0.199	0.207	407.101	0.141	0.168	0.165	347.918	0.113	0.156	0.168

Notes: See Table B1.

Table B3. Monte Carlo Results: Estimates in the Common Effect Model ($\rho_0\sigma_0$ = -0.50; ρ_{01} = 1)

	Homoskedastic Error in Treatment Equation								Heteroskedastic Error in Treatment Equation							
	ATE				ATT				ATE				ATT			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
I. Normally Distributed Errors																
τ_{HI}	1.308	0.884	0.884	0.884	1.267	0.894	0.894	0.894	1.300	0.844	0.876	0.876	1.258	0.846	0.878	0.878
$\tau_{MB,0.05}$	1.098	0.801	0.801	0.800	0.959	0.799	0.800	0.799	1.033	0.825	0.801	0.801	1.400	0.819	0.802	0.801
$\tau_{MB,0.25}$	1.150	0.808	0.808	0.808	1.266	0.808	0.808	0.808	1.075	0.827	0.807	0.807	1.438	0.828	0.807	0.807
$\tau_{MB,EE,0.05}$	0.893	0.871	0.873	0.873	1.192	0.844	0.854	0.850	1.141	0.804	0.804	0.805	1.132	0.821	0.816	0.814
$\tau_{MB,EE,0.25}$	1.135	0.842	0.844	0.845	1.128	0.827	0.837	0.834	1.219	0.828	0.808	0.808	1.169	0.830	0.818	0.819
τ_{KV}	3.309	0.053	0.057	0.058	3.309	0.053	0.057	0.058	3.239	0.043	0.055	0.055	3.239	0.043	0.055	0.055
τ_{BVN}	13.982	0.035	0.038	0.038	9.453	0.221	0.221	0.221	8.065	0.156	0.065	0.065	5.751	0.325	0.247	0.248
$\tau_{MB\text{-}BC,0.05}$	13.527	0.038	0.039	0.039	12.180	0.036	0.038	0.037	7.631	0.180	0.067	0.067	7.400	0.188	0.067	0.066
$\tau_{MB\text{-}BC,0.25}$	13.580	0.038	0.040	0.040	12.487	0.037	0.039	0.039	7.672	0.182	0.070	0.071	7.438	0.197	0.070	0.070
$\tau_{BVN,\ EE}$	193.976	0.219	0.217	0.215	101.757	0.263	0.256	0.257	362.288	0.186	0.118	0.119	221.462	0.219	0.201	0.202
$\tau_{MB\text{-}BC,EE,0.05}$	193.871	0.288	0.293	0.260	146.741	0.225	0.236	0.220	362.089	0.228	0.129	0.129	309.975	0.165	0.131	0.135
$\tau_{MB\text{-}BC,EE,0.25}$	193.630	0.308	0.312	0.280	146.803	0.220	0.231	0.218	362.166	0.198	0.132	0.132	310.013	0.150	0.132	0.136
II. Asymmetric, Non-Normally Distributed Errors																
τ_{HI}	1.144	0.624	0.624	0.624	1.083	0.485	0.485	0.485	1.148	0.638	0.668	0.668	1.088	0.510	0.531	0.531
$\tau_{MB,0.05}$	1.249	0.210	0.195	0.195	1.089	0.779	0.779	0.779	1.669	0.105	0.403	0.403	1.433	0.800	0.777	0.777
$\tau_{MB,0.25}$	1.325	0.369	0.369	0.369	1.237	0.715	0.715	0.715	1.503	0.381	0.433	0.433	1.418	0.731	0.714	0.714
$\tau_{MB,EE,0.05}$	0.703	0.238	0.241	0.241	0.686	0.210	0.209	0.211	1.666	0.113	0.289	0.290	0.813	0.104	0.324	0.287
$\tau_{MB,EE,0.25}$	0.919	0.374	0.373	0.373	0.754	0.353	0.352	0.353	1.502	0.382	0.421	0.421	0.958	0.362	0.420	0.410
τ_{KV}	3.306	0.052	0.060	0.060	3.306	0.052	0.060	0.060	3.233	0.033	0.046	0.046	3.233	0.033	0.046	0.046
τ_{BVN}	7.576	0.288	0.144	0.145	5.638	0.573	0.492	0.493	2.934	0.311	0.376	0.377	2.644	0.574	0.644	0.645
$\tau_{MB\text{-}BC,0.05}$	7.352	0.207	0.146	0.147	7.328	0.848	0.773	0.774	2.892	0.105	0.403	0.403	3.501	0.866	0.959	0.960
$\tau_{MB\text{-}BC,0.25}$	7.431	0.366	0.315	0.316	7.475	0.784	0.709	0.710	2.731	0.381	0.433	0.433	3.487	0.798	0.896	0.897
$\tau_{BVN,\ EE}$	248.396	0.215	0.220	0.218	133.438	0.224	0.221	0.224	308.049	0.158	0.162	0.164	189.241	0.265	0.287	0.289
$\tau_{MB\text{-}BC,EE,0.05}$	248.219	0.230	0.229	0.230	192.306	0.202	0.197	0.199	305.633	0.114	0.288	0.289	265.048	0.104	0.297	0.286
$\tau_{MB\text{-}BC,EE,0.25}$	248.003	0.359	0.354	0.356	192.237	0.340	0.335	0.336	305.469	0.366	0.420	0.420	265.192	0.351	0.428	0.409

Notes: See Table B1.

Table B4. Monte Carlo Results: Estimates in the Heterogeneous Effect Model ($\rho_0\sigma_0 = 0$; $\rho_{01} = 0.50$; $\rho_\delta\sigma_\delta = 0$)

	Homoskedastic Error in Treatment Equation								Heteroskedastic Error in Treatment Equation							
	ATE				ATT				ATE				ATT			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
I. Normally Distributed Errors																
τ_{HI}	0.614	0.006	0.006	0.006	0.560	0.007	0.007	0.007	0.616	0.014	0.010	0.010	0.566	0.009	0.008	0.008
$\tau_{MB,0.05}$	0.311	0.016	0.015	0.016	0.197	0.015	0.014	0.014	0.228	0.047	0.025	0.025	0.729	0.029	0.015	0.015
$\tau_{MB,0.25}$	0.380	0.008	0.008	0.007	0.590	0.006	0.006	0.006	0.282	0.023	0.013	0.013	0.772	0.025	0.007	0.007
$\tau_{MB,EE,0.05}$	0.278	0.017	0.017	0.017	0.424	0.016	0.017	0.018	0.397	0.095	0.027	0.027	0.341	0.095	0.028	0.028
$\tau_{MB,EE,0.25}$	0.534	0.010	0.009	0.009	0.330	0.009	0.010	0.009	0.537	0.023	0.014	0.013	0.432	0.031	0.013	0.013
τ_{KV}	3.309	0.049	0.048	0.049	3.309	0.049	0.048	0.049	3.238	0.035	0.042	0.042	3.238	0.035	0.042	0.042
τ_{BVN}	17.365	0.034	0.038	0.038	11.370	0.025	0.028	0.028	10.058	0.033	0.032	0.032	6.797	0.026	0.025	0.025
$\tau_{MB\text{-}BC,0.05}$	16.744	0.036	0.039	0.040	15.025	0.036	0.039	0.039	9.450	0.057	0.040	0.040	9.068	0.048	0.035	0.035
$\tau_{MB\text{-}BC,0.25}$	16.813	0.034	0.037	0.038	15.419	0.034	0.037	0.037	9.505	0.038	0.032	0.033	9.111	0.042	0.032	0.033
$\tau_{BVN,\ EE}$	166.554	0.161	0.169	0.170	84.849	0.108	0.120	0.121	489.627	0.096	0.118	0.117	296.878	0.069	0.103	0.102
$\tau_{MB\text{-}BC,EE,0.05}$	165.844	0.160	0.206	0.207	121.848	0.140	0.148	0.151	489.410	0.117	0.149	0.149	416.850	0.128	0.121	0.120
$\tau_{MB\text{-}BC,EE,0.25}$	165.567	0.158	0.203	0.204	121.942	0.139	0.147	0.150	489.550	0.080	0.142	0.141	416.943	0.092	0.114	0.113
II. Asymmetric, Non-Normally Distributed Errors																
τ_{HI}	0.613	0.006	0.006	0.006	0.560	0.008	0.007	0.007	0.615	0.015	0.009	0.009	0.565	0.010	0.007	0.007
$\tau_{MB,0.05}$	0.313	0.016	0.017	0.016	0.197	0.017	0.016	0.016	0.226	0.083	0.020	0.021	0.728	0.027	0.016	0.016
$\tau_{MB,0.25}$	0.377	0.009	0.010	0.010	0.586	0.009	0.008	0.009	0.283	0.021	0.010	0.009	0.770	0.024	0.008	0.008
$\tau_{MB,EE,0.05}$	0.299	0.020	0.020	0.021	0.416	0.018	0.018	0.019	0.391	0.104	0.033	0.033	0.344	0.099	0.028	0.028
$\tau_{MB,EE,0.25}$	0.552	0.011	0.010	0.011	0.330	0.009	0.009	0.009	0.535	0.020	0.015	0.015	0.430	0.027	0.013	0.012
τ_{KV}	3.309	0.056	0.074	0.074	3.309	0.056	0.074	0.074	3.236	0.041	0.059	0.059	3.236	0.041	0.059	0.059
τ_{BVN}	17.273	0.043	0.047	0.047	11.312	0.033	0.035	0.035	10.112	0.035	0.049	0.049	6.832	0.026	0.037	0.037
$\tau_{MB\text{-}BC,0.05}$	16.654	0.045	0.048	0.048	14.946	0.050	0.052	0.051	9.501	0.088	0.052	0.053	9.114	0.045	0.053	0.053
$\tau_{MB\text{-}BC,0.25}$	16.718	0.042	0.046	0.046	15.335	0.045	0.047	0.047	9.559	0.041	0.048	0.047	9.157	0.042	0.047	0.047
$\tau_{BVN,\ EE}$	162.615	0.206	0.216	0.215	82.640	0.140	0.152	0.151	485.568	0.102	0.112	0.112	294.464	0.069	0.095	0.095
$\tau_{MB\text{-}BC,EE,0.05}$	161.783	0.191	0.214	0.200	118.670	0.183	0.191	0.197	485.350	0.135	0.153	0.153	413.432	0.116	0.121	0.120
$\tau_{MB\text{-}BC,EE,0.25}$	161.510	0.185	0.210	0.196	118.756	0.178	0.186	0.192	485.494	0.101	0.144	0.144	413.518	0.092	0.114	0.112

Notes: See Table B1.

Table B5. Monte Carlo Results: Estimates in the Heterogeneous Effect Model ($\rho_0\sigma_0$ = -0.20; ρ_{01} = 0.50; $\rho_\delta\sigma_\delta$ = -0.10)

	Homoskedastic Error in Treatment Equation								Heteroskedastic Error in Treatment Equation							
	ATE				ATT				ATE				ATT			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
I. Normally Distributed Errors																
τ_{HI}	0.936	0.413	0.413	0.413	0.841	0.359	0.359	0.359	0.933	0.390	0.418	0.418	0.841	0.337	0.360	0.360
$\tau_{MB,0.05}$	0.628	0.383	0.383	0.383	0.515	0.309	0.310	0.309	0.655	0.411	0.382	0.383	0.985	0.335	0.310	0.310
$\tau_{MB,0.25}$	0.724	0.382	0.382	0.381	0.842	0.307	0.307	0.307	0.707	0.402	0.381	0.381	1.019	0.330	0.307	0.307
$\tau_{MB,EE,0.05}$	0.489	0.407	0.414	0.411	0.672	0.331	0.328	0.326	0.886	0.384	0.402	0.400	0.594	0.346	0.325	0.322
$\tau_{MB,EE,0.25}$	0.733	0.392	0.395	0.394	0.590	0.310	0.308	0.307	1.040	0.393	0.391	0.390	0.679	0.327	0.306	0.307
τ_{KV}	3.306	0.039	0.060	0.060	3.263	0.060	0.062	0.062	3.230	0.048	0.061	0.061	3.188	0.084	0.095	0.094
τ_{BVN}	14.491	0.036	0.046	0.046	9.661	0.095	0.094	0.094	8.278	0.085	0.045	0.045	5.760	0.144	0.104	0.104
$\tau_{MB\text{-}BC,0.05}$	13.908	0.042	0.049	0.050	12.606	0.039	0.048	0.048	7.806	0.119	0.049	0.049	7.558	0.103	0.040	0.040
$\tau_{MB\text{-}BC,0.25}$	14.004	0.037	0.046	0.046	12.933	0.036	0.046	0.046	7.859	0.109	0.046	0.046	7.594	0.097	0.037	0.038
$\tau_{BVN,\ EE}$	176.333	0.218	0.218	0.216	91.110	0.185	0.183	0.182	413.622	0.162	0.145	0.145	251.899	0.115	0.162	0.163
$\tau_{MB\text{-}BC,EE,0.05}$	176.203	0.240	0.275	0.270	131.159	0.223	0.216	0.220	413.388	0.177	0.166	0.136	353.104	0.187	0.209	0.209
$\tau_{MB\text{-}BC,EE,0.25}$	175.960	0.248	0.289	0.281	131.240	0.206	0.200	0.203	413.540	0.165	0.175	0.134	353.189	0.160	0.188	0.193
II. Asymmetric, Non-Normally Distributed Errors																
τ_{HI}	0.904	0.376	0.376	0.376	0.802	0.281	0.281	0.281	0.904	0.362	0.389	0.389	0.817	0.285	0.304	0.304
$\tau_{MB,0.05}$	0.536	0.284	0.285	0.285	0.606	0.364	0.364	0.364	0.858	0.106	0.302	0.301	1.051	0.383	0.357	0.357
$\tau_{MB,0.25}$	0.644	0.297	0.297	0.297	0.890	0.344	0.344	0.344	0.870	0.268	0.310	0.310	1.068	0.359	0.337	0.337
$\tau_{MB,EE,0.05}$	0.456	0.304	0.307	0.309	0.606	0.269	0.297	0.305	1.080	0.118	0.312	0.301	0.590	0.269	0.313	0.332
$\tau_{MB,EE,0.25}$	0.691	0.312	0.315	0.316	0.561	0.271	0.282	0.284	1.183	0.268	0.315	0.310	0.688	0.264	0.287	0.294
τ_{KV}	3.276	0.137	0.122	0.121	3.229	0.181	0.164	0.163	3.204	0.107	0.123	0.122	3.171	0.138	0.154	0.153
τ_{BVN}	12.671	0.091	0.142	0.142	8.570	0.069	0.045	0.045	6.696	0.033	0.035	0.035	4.807	0.137	0.138	0.139
$\tau_{MB\text{-}BC,0.05}$	12.031	0.097	0.141	0.140	11.257	0.108	0.083	0.083	6.447	0.171	0.054	0.054	6.381	0.204	0.185	0.186
$\tau_{MB\text{-}BC,0.25}$	12.140	0.087	0.129	0.129	11.541	0.090	0.065	0.066	6.462	0.036	0.059	0.060	6.399	0.181	0.166	0.166
$\tau_{BVN,\ EE}$	208.984	0.259	0.263	0.264	109.793	0.182	0.187	0.188	394.018	0.146	0.106	0.106	240.460	0.116	0.134	0.136
$\tau_{MB\text{-}BC,EE,0.05}$	208.833	0.235	0.339	0.339	158.126	0.233	0.263	0.257	393.307	0.250	0.184	0.132	337.104	0.174	0.167	0.191
$\tau_{MB\text{-}BC,EE,0.25}$	208.599	0.230	0.350	0.349	158.169	0.225	0.246	0.234	393.411	0.144	0.204	0.137	337.203	0.126	0.133	0.143

Notes: See Table B1.

Table B6. Monte Carlo Results: Estimates in the Heterogeneous Effect Model ($\rho_0\sigma_0$ = -0.40; ρ_{01} = 0.50; $\rho_\delta\sigma_\delta$ = -0.10)

	Homoskedastic Error in Treatment Equation								Heteroskedastic Error in Treatment Equation							
	ATE				ATT				ATE				ATT			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
I. Normally Distributed Errors																
τ_{HI}	1.214	0.767	0.767	0.767	1.125	0.717	0.716	0.716	1.207	0.733	0.765	0.765	1.119	0.678	0.708	0.708
$\tau_{MB,0.05}$	0.936	0.710	0.710	0.710	0.830	0.646	0.646	0.645	1.009	0.734	0.707	0.707	1.265	0.667	0.646	0.646
$\tau_{MB,0.25}$	1.027	0.710	0.711	0.711	1.118	0.642	0.642	0.642	1.049	0.736	0.709	0.709	1.297	0.661	0.642	0.642
$\tau_{MB,EE,0.05}$	0.802	0.766	0.772	0.772	0.987	0.679	0.682	0.683	1.107	0.719	0.713	0.712	0.924	0.669	0.649	0.649
$\tau_{MB,EE,0.25}$	1.027	0.737	0.738	0.739	0.919	0.652	0.651	0.653	1.222	0.733	0.713	0.712	0.984	0.660	0.641	0.641
τ_{KV}	3.308	0.050	0.056	0.055	3.265	0.079	0.070	0.069	3.231	0.061	0.072	0.071	3.189	0.097	0.107	0.106
τ_{BVN}	13.085	0.040	0.046	0.046	8.860	0.177	0.177	0.177	7.275	0.146	0.061	0.061	5.212	0.276	0.201	0.201
$\tau_{MB\text{-}BC,0.05}$	12.563	0.043	0.050	0.050	11.436	0.040	0.046	0.045	6.906	0.175	0.063	0.062	6.731	0.182	0.068	0.067
$\tau_{MB\text{-}BC,0.25}$	12.654	0.043	0.049	0.048	11.724	0.040	0.045	0.045	6.946	0.176	0.062	0.062	6.763	0.176	0.064	0.064
$\tau_{BVN,\ EE}$	190.273	0.211	0.209	0.209	99.565	0.208	0.214	0.214	369.806	0.186	0.126	0.126	225.825	0.174	0.186	0.186
$\tau_{MB\text{-}BC,EE,0.05}$	190.160	0.299	0.283	0.292	143.592	0.233	0.248	0.241	369.558	0.205	0.120	0.121	316.236	0.179	0.194	0.195
$\tau_{MB\text{-}BC,EE,0.25}$	189.935	0.324	0.303	0.317	143.660	0.235	0.243	0.236	369.672	0.186	0.117	0.118	316.295	0.152	0.180	0.181
II. Asymmetric, Non-Normally Distributed Errors																
τ_{HI}	1.093	0.602	0.602	0.602	0.994	0.441	0.441	0.441	1.095	0.603	0.634	0.634	1.010	0.461	0.484	0.484
$\tau_{MB,0.05}$	1.131	0.282	0.282	0.282	0.998	0.691	0.691	0.691	1.609	0.145	0.338	0.339	1.358	0.722	0.698	0.699
$\tau_{MB,0.25}$	1.258	0.387	0.387	0.386	1.162	0.634	0.634	0.634	1.441	0.379	0.422	0.422	1.353	0.660	0.641	0.641
$\tau_{MB,EE,0.05}$	0.653	0.313	0.312	0.312	0.672	0.285	0.289	0.288	1.608	0.145	0.337	0.338	0.789	0.134	0.416	0.332
$\tau_{MB,EE,0.25}$	0.868	0.400	0.398	0.398	0.718	0.372	0.374	0.374	1.444	0.378	0.422	0.422	0.919	0.361	0.437	0.413
τ_{KV}	3.248	0.227	0.202	0.201	3.211	0.264	0.238	0.237	3.177	0.150	0.164	0.164	3.152	0.175	0.189	0.188
τ_{BVN}	7.656	0.067	0.199	0.199	5.628	0.259	0.183	0.183	3.075	0.074	0.125	0.125	2.684	0.343	0.397	0.397
$\tau_{MB\text{-}BC,0.05}$	7.369	0.119	0.231	0.231	7.339	0.456	0.383	0.383	3.091	0.112	0.224	0.225	3.580	0.583	0.645	0.646
$\tau_{MB\text{-}BC,0.25}$	7.500	0.048	0.133	0.133	7.504	0.400	0.326	0.326	2.927	0.136	0.307	0.307	3.576	0.520	0.588	0.588
$\tau_{BVN,\ EE}$	256.032	0.270	0.260	0.260	137.411	0.215	0.205	0.208	296.996	0.144	0.118	0.120	183.244	0.209	0.216	0.218
$\tau_{MB\text{-}BC,EE,0.05}$	255.865	0.252	0.248	0.248	197.951	0.222	0.228	0.228	294.303	0.138	0.286	0.288	256.301	0.125	0.283	0.284
$\tau_{MB\text{-}BC,EE,0.25}$	255.650	0.296	0.295	0.295	197.904	0.267	0.275	0.275	294.140	0.266	0.368	0.369	256.431	0.263	0.361	0.363

Notes: See Table B1.

Table C1. Effect of SBP Participation on BMI Growth

	Third Grade Outcome			Fifth Grade Outcome		
	Specification (1)	Specification (2)	Specification (3)	Specification (1)	Specification (2)	Specification (3)
A. ATE						
τ_{OLS}	0.008	0.008	0.008	0.441	0.524	0.523
	[0.004, 0.011]	[0.004, 0.012]	[0.005, 0.013]	[-0.596, 1.440]	[-0.397, 1.421]	[-0.356, 1.390]
τ_{HI}	0.011	0.011	0.011	1.108	0.568	1.031
	[0.006, 0.016]	[0.006, 0.016]	[0.007, 0.016]	[-0.250, 2.117]	[-0.638, 2.016]	[-0.278, 2.096]
$\tau_{MB,0.05}$	0.007	0.012	0.016	1.768	-0.325	0.133
	[-0.007, 0.020]	[-0.004, 0.020]	[-0.005, 0.023]	[-2.505, 3.297]	[-2.615, 3.655]	[-2.805, 4.000]
$\tau_{MB,0.25}$	0.007	0.010	0.010	0.855	0.263	0.631
	[0.001, 0.012]	[0.002, 0.013]	[0.003, 0.015]	[-0.916, 1.772]	[-1.037, 2.038]	[-0.647, 1.951]
τ_{KV}	-0.041	-0.037	-0.037	-5.168	-0.287	-3.029
	[-0.064, -0.014]	[-0.060, -0.006]	[-0.063, -0.008]	[-9.681, -0.186]	[-6.138, 5.347]	[-8.575, 3.430]
τ_{BVN}	-0.045	-0.028	-0.032	-4.882	-1.301	-3.960
	[-0.069, -0.022]	[-0.054, -0.002]	[-0.052, -0.007]	[-10.226, 0.329]	[-6.913, 3.918]	[-8.903, 2.686]
$\tau_{MB\text{-}BC,0.05}$	-0.044	-0.022	-0.022	-3.410	-1.682	-4.060
	[-0.075, -0.015]	[-0.054, 0.005]	[-0.053, -0.001]	[-9.916, 0.388]	[-7.331, 5.246]	[-9.292, 3.071]
$\tau_{MB\text{-}BC,0.25}$	-0.044	-0.024	-0.028	-4.323	-1.094	-3.563
	[-0.069, -0.019]	[-0.053, 0.003]	[-0.049, 0.002]	[-9.814, 0.670]	[-6.115, 3.779]	[-7.975, 2.609]
P*	0.415	0.514	0.546	0.392	0.807	0.510
	[0.263, 0.704]	[0.216, 0.894]	[0.317, 0.925]	[0.080, 0.868]	[0.030, 0.923]	[0.057, 0.880]
B. ATT						
τ_{OLS}	0.008	0.008	0.008	0.441	0.524	0.523
	[0.004, 0.011]	[0.004, 0.012]	[0.005, 0.013]	[-0.596, 1.440]	[-0.397, 1.421]	[-0.356, 1.390]
τ_{HI}	0.009	0.009	0.008	0.558	0.431	0.555
	[0.004, 0.013]	[0.005, 0.014]	[0.004, 0.013]	[-0.418, 1.487]	[-0.578, 1.376]	[-0.393, 1.595]
$\tau_{MB,0.05}$	0.012	0.013	0.010	1.086	0.641	0.455
	[0.001, 0.024]	[0.001, 0.025]	[-0.001, 0.026]	[-1.907, 3.341]	[-1.974, 3.028]	[-2.412, 2.968]
$\tau_{MB,0.25}$	0.012	0.012	0.010	0.532	0.273	0.145
	[0.005, 0.016]	[0.006, 0.016]	[0.005, 0.016]	[-0.841, 1.884]	[-0.781, 1.523]	[-0.876, 1.479]
τ_{KV}	-0.041	-0.037	-0.037	-5.168	-0.287	-3.029
	[-0.064, -0.014]	[-0.060, -0.006]	[-0.063, -0.008]	[-9.681, -0.186]	[-6.138, 5.347]	[-8.575, 3.430]
τ_{BVN}	-0.014	-0.006	-0.007	-1.703	0.039	-1.152
	[-0.028, -0.003]	[-0.021, 0.007]	[-0.017, 0.007]	[-4.458, 0.988]	[-2.598, 2.718]	[-3.587, 2.258]
$\tau_{MB\text{-}BC,0.05}$	-0.043	-0.021	-0.026	-4.652	-0.404	-3.699
	[-0.077, -0.013]	[-0.054, 0.012]	[-0.050, 0.012]	[-11.959, 2.261]	[-7.177, 6.872]	[-10.352, 4.554]
$\tau_{MB\text{-}BC,0.25}$	-0.043	-0.022	-0.026	-5.206	-0.772	-4.008
	[-0.076, -0.016]	[-0.056, 0.007]	[-0.050, 0.007]	[-11.844, 1.424]	[-6.610, 5.897]	[-10.397, 4.814]
P*	0.500	0.500	0.500	0.500	0.500	0.500

Notes: Treatment is defined as participation in SBP in first and third grades (first, third, and fifth grades) in Columns 1-3 (4-6). See Table 4 for more details.

Table C2. Effect of SBP Participation on the Probability of Being Overweight

	Third Grade Outcome			Fifth Grade Outcome		
	Specification (1)	Specification (2)	Specification (3)	Specification (1)	Specification (2)	Specification (3)
A. ATE						
τ_{OLS}	0.023	0.023	0.024	0.020	0.021	0.020
	[0.002, 0.045]	[0.001, 0.046]	[0.002, 0.044]	[-0.003, 0.042]	[-0.005, 0.045]	[-0.006, 0.045]
τ_{HI}	0.038	0.037	0.037	0.029	0.035	0.031
	[0.009, 0.071]	[0.014, 0.066]	[0.008, 0.063]	[0.003, 0.063]	[0.005, 0.072]	[-0.003, 0.067]
$\tau_{MB,0.05}$	-0.025	-0.036	-0.005	0.033	0.087	0.100
	[-0.080, 0.063]	[-0.060, 0.078]	[-0.064, 0.081]	[-0.073, 0.123]	[-0.065, 0.122]	[-0.075, 0.141]
$\tau_{MB,0.25}$	-0.001	0.016	0.017	0.033	0.053	0.036
	[-0.028, 0.037]	[-0.017, 0.048]	[-0.019, 0.041]	[-0.019, 0.067]	[-0.020, 0.068]	[-0.012, 0.072]
τ_{KV}	-0.175	-0.176	-0.159	-0.219	-0.223	-0.228
	[-0.306, -0.032]	[-0.292, -0.006]	[-0.272, 0.020]	[-0.372, -0.079]	[-0.367, -0.057]	[-0.339, -0.048]
τ_{BVN}	-0.269	-0.227	-0.213	-0.282	-0.239	-0.255
	[-0.389, -0.148]	[-0.340, -0.084]	[-0.319, -0.059]	[-0.408, -0.146]	[-0.372, -0.079]	[-0.363, -0.099]
$\tau_{MB\text{-}BC,0.05}$	-0.309	-0.279	-0.233	-0.263	-0.165	-0.167
	[-0.430, -0.120]	[-0.352, -0.080]	[-0.349, -0.030]	[-0.392, -0.126]	[-0.369, -0.055]	[-0.373, -0.054]
$\tau_{MB\text{-}BC,0.25}$	-0.285	-0.227	-0.212	-0.263	-0.199	-0.230
	[-0.410, -0.139]	[-0.339, -0.081]	[-0.313, -0.052]	[-0.389, -0.133]	[-0.356, -0.048]	[-0.347, -0.063]
P*	0.343	0.343	0.392	0.279	0.307	0.346
	[0.228, 0.639]	[0.196, 0.683]	[0.219, 0.760]	[0.139, 0.470]	[0.155, 0.676]	[0.167, 0.672]
B. ATT						
τ_{OLS}	0.023	0.023	0.024	0.020	0.021	0.020
	[0.002, 0.045]	[0.001, 0.046]	[0.002, 0.044]	[-0.003, 0.042]	[-0.005, 0.045]	[-0.006, 0.045]
τ_{HI}	0.030	0.030	0.024	0.035	0.028	0.024
	[0.004, 0.054]	[0.008, 0.053]	[0.002, 0.047]	[0.007, 0.061]	[0.000, 0.057]	[-0.006, 0.051]
$\tau_{MB,0.05}$	0.024	0.034	0.017	0.001	0.025	0.001
	[-0.034, 0.085]	[-0.031, 0.096]	[-0.033, 0.089]	[-0.049, 0.077]	[-0.075, 0.071]	[-0.079, 0.056]
$\tau_{MB,0.25}$	0.032	0.037	0.032	0.029	0.030	0.037
	[-0.005, 0.050]	[0.005, 0.063]	[-0.000, 0.057]	[0.003, 0.064]	[-0.001, 0.070]	[-0.002, 0.067]
τ_{KV}	-0.175	-0.176	-0.159	-0.219	-0.223	-0.228
	[-0.306, -0.032]	[-0.292, -0.006]	[-0.272, 0.020]	[-0.372, -0.079]	[-0.367, -0.057]	[-0.339, -0.048]
τ_{BVN}	-0.108	-0.091	-0.081	-0.116	-0.095	-0.099
	[-0.167, -0.045]	[-0.150, -0.019]	[-0.138, -0.004]	[-0.180, -0.053]	[-0.164, -0.007]	[-0.158, -0.017]
$\tau_{MB\text{-}BC,0.05}$	-0.307	-0.249	-0.236	-0.374	-0.282	-0.309
	[-0.454, -0.138]	[-0.415, -0.073]	[-0.371, -0.034]	[-0.539, -0.175]	[-0.484, -0.079]	[-0.465, -0.119]
$\tau_{MB\text{-}BC,0.25}$	-0.299	-0.245	-0.221	-0.345	-0.277	-0.273
	[-0.446, -0.149]	[-0.395, -0.080]	[-0.347, -0.031]	[-0.493, -0.180]	[-0.448, -0.060]	[-0.427, -0.058]
P*	0.500	0.500	0.500	0.500	0.500	0.500

Notes: Overweight is defined as BMI above the 85th percentile. For other details, see Table C1.

Table C3. Effect of SBP Participation on the Probability of Being Obese

	Third Grade Outcome			Fifth Grade Outcome		
	Specification (1)	Specification (2)	Specification (3)	Specification (1)	Specification (2)	Specification (3)
A. ATE						
τ_{OLS}	0.013	0.014	0.015	0.018	0.020	0.019
	[-0.002, 0.029]	[-0.002, 0.031]	[-0.003, 0.035]	[-0.005, 0.040]	[-0.000, 0.041]	[-0.004, 0.042]
τ_{HI}	0.024	0.030	0.026	0.037	0.044	0.041
	[0.004, 0.052]	[0.010, 0.060]	[0.005, 0.049]	[0.010, 0.067]	[0.017, 0.075]	[0.013, 0.073]
$\tau_{MB,0.05}$	0.011	0.003	0.023	0.033	-0.004	-0.008
	[-0.058, 0.060]	[-0.041, 0.066]	[-0.049, 0.071]	[-0.049, 0.092]	[-0.059, 0.112]	[-0.063, 0.096]
$\tau_{MB,0.25}$	0.008	0.010	0.022	0.022	0.028	0.019
	[-0.024, 0.032]	[-0.014, 0.034]	[-0.019, 0.043]	[-0.014, 0.053]	[-0.008, 0.058]	[-0.019, 0.044]
τ_{KV}	-0.176	-0.141	-0.134	-0.235	-0.197	-0.194
	[-0.277, -0.061]	[-0.230, -0.003]	[-0.214, -0.007]	[-0.340, -0.121]	[-0.294, -0.068]	[-0.298, -0.050]
τ_{BVN}	-0.267	-0.196	-0.196	-0.274	-0.206	-0.236
	[-0.363, -0.161]	[-0.268, -0.083]	[-0.274, -0.060]	[-0.384, -0.163]	[-0.309, -0.080]	[-0.322, -0.077]
$\tau_{MB\text{-}BC,0.05}$	-0.260	-0.199	-0.179	-0.249	-0.215	-0.246
	[-0.373, -0.137]	[-0.288, -0.065]	[-0.282, -0.056]	[-0.388, -0.136]	[-0.309, -0.023]	[-0.329, -0.055]
$\tau_{MB\text{-}BC,0.25}$	-0.263	-0.193	-0.180	-0.260	-0.183	-0.219
	[-0.362, -0.157]	[-0.272, -0.066]	[-0.268, -0.049]	[-0.369, -0.139]	[-0.300, -0.017]	[-0.302, -0.043]
P*	0.388	0.403	0.427	0.434	0.526	0.514
	[0.237, 0.570]	[0.234, 0.754]	[0.219, 0.766]	[0.272, 0.707]	[0.296, 0.914]	[0.303, 0.899]
B. ATT						
τ_{OLS}	0.013	0.014	0.015	0.018	0.020	0.019
	[-0.002, 0.029]	[-0.002, 0.031]	[-0.003, 0.035]	[-0.005, 0.040]	[-0.000, 0.041]	[-0.004, 0.042]
τ_{HI}	0.018	0.017	0.013	0.029	0.025	0.021
	[0.001, 0.037]	[-0.000, 0.036]	[-0.006, 0.033]	[0.003, 0.054]	[0.002, 0.048]	[-0.006, 0.045]
$\tau_{MB,0.05}$	0.023	0.026	0.021	-0.010	0.005	-0.013
	[-0.032, 0.065]	[-0.028, 0.074]	[-0.029, 0.071]	[-0.061, 0.075]	[-0.057, 0.067]	[-0.065, 0.062]
$\tau_{MB,0.25}$	0.022	0.023	0.023	0.022	0.021	0.023
	[-0.002, 0.039]	[-0.003, 0.043]	[-0.001, 0.050]	[-0.001, 0.056]	[-0.007, 0.055]	[-0.008, 0.051]
τ_{KV}	-0.176	-0.141	-0.134	-0.235	-0.197	-0.194
	[-0.277, -0.061]	[-0.230, -0.003]	[-0.214, -0.007]	[-0.340, -0.121]	[-0.294, -0.068]	[-0.298, -0.050]
τ_{BVN}	-0.109	-0.077	-0.075	-0.095	-0.063	-0.076
	[-0.158, -0.060]	[-0.117, -0.018]	[-0.123, -0.006]	[-0.153, -0.040]	[-0.118, 0.008]	[-0.124, -0.001]
$\tau_{MB\text{-}BC,0.05}$	-0.278	-0.195	-0.194	-0.309	-0.202	-0.248
	[-0.410, -0.149]	[-0.295, -0.040]	[-0.315, -0.025]	[-0.449, -0.147]	[-0.335, -0.034]	[-0.347, -0.035]
$\tau_{MB\text{-}BC,0.25}$	-0.279	-0.198	-0.193	-0.277	-0.185	-0.212
	[-0.391, -0.157]	[-0.294, -0.056]	[-0.302, -0.037]	[-0.409, -0.119]	[-0.332, -0.007]	[-0.317, -0.029]
P*	0.500	0.500	0.500	0.500	0.500	0.500

Notes: Obese is defined as BMI above the 95th percentile. For other details, see Table C1.

Table C4. Effect of SBP Participation on the Probability of Being Underweight

	Third Grade Outcome			Fifth Grade Outcome		
	Specification (1)	Specification (2)	Specification (3)	Specification (1)	Specification (2)	Specification (3)
A. ATE						
τ_{OLS}	-0.020	-0.021	-0.020	-0.017	-0.018	-0.017
	[-0.031, -0.009]	[-0.030, -0.009]	[-0.030, -0.008]	[-0.029, -0.005]	[-0.030, -0.005]	[-0.029, -0.005]
τ_{HI}	-0.026	-0.025	-0.025	-0.032	-0.027	-0.025
	[-0.040, -0.015]	[-0.036, -0.012]	[-0.037, -0.010]	[-0.044, -0.016]	[-0.039, -0.012]	[-0.036, -0.010]
$\tau_{MB,0.05}$	-0.016	-0.018	-0.021	0.006	0.000	-0.004
	[-0.058, 0.018]	[-0.064, 0.024]	[-0.065, 0.026]	[-0.053, 0.030]	[-0.058, 0.030]	[-0.059, 0.031]
$\tau_{MB,0.25}$	-0.017	-0.016	-0.016	-0.010	-0.011	-0.011
	[-0.038, 0.001]	[-0.042, -0.001]	[-0.040, -0.000]	[-0.030, 0.003]	[-0.039, 0.006]	[-0.041, 0.006]
τ_{KV}	-0.007	-0.018	0.003	0.091	0.036	0.054
	[-0.085, 0.078]	[-0.107, 0.073]	[-0.095, 0.100]	[-0.003, 0.200]	[-0.090, 0.134]	[-0.066, 0.145]
τ_{BVN}	0.018	0.002	0.016	0.115	0.048	0.054
	[-0.058, 0.107]	[-0.098, 0.087]	[-0.082, 0.104]	[0.019, 0.222]	[-0.061, 0.147]	[-0.056, 0.146]
$\tau_{MB\text{-}BC,0.05}$	0.019	-0.003	0.010	0.129	0.054	0.055
	[-0.072, 0.118]	[-0.095, 0.102]	[-0.088, 0.110]	[0.003, 0.210]	[-0.067, 0.150]	[-0.066, 0.149]
$\tau_{MB\text{-}BC,0.25}$	0.017	0.000	0.015	0.113	0.042	0.048
	[-0.061, 0.105]	[-0.096, 0.096]	[-0.079, 0.104]	[-0.004, 0.216]	[-0.061, 0.143]	[-0.056, 0.137]
P*	0.679	0.880	0.711	0.577	0.793	0.747
	[0.034, 0.961]	[0.028, 0.925]	[0.038, 0.897]	[0.350, 0.948]	[0.056, 0.977]	[0.032, 0.971]
B. ATT						
τ_{OLS}	-0.020	-0.021	-0.020	-0.017	-0.018	-0.017
	[-0.031, -0.009]	[-0.030, -0.009]	[-0.030, -0.008]	[-0.029, -0.005]	[-0.030, -0.005]	[-0.029, -0.005]
τ_{HI}	-0.020	-0.017	-0.017	-0.020	-0.013	-0.011
	[-0.034, -0.006]	[-0.030, -0.003]	[-0.031, -0.005]	[-0.039, -0.004]	[-0.028, 0.002]	[-0.025, 0.002]
$\tau_{MB,0.05}$	-0.014	-0.024	-0.026	-0.019	-0.042	-0.042
	[-0.062, 0.019]	[-0.060, 0.024]	[-0.057, 0.021]	[-0.062, 0.020]	[-0.059, 0.025]	[-0.062, 0.018]
$\tau_{MB,0.25}$	-0.015	-0.018	-0.015	-0.009	-0.012	-0.015
	[-0.030, 0.000]	[-0.029, 0.002]	[-0.030, 0.002]	[-0.028, 0.006]	[-0.031, 0.008]	[-0.029, 0.007]
τ_{KV}	-0.007	-0.018	0.003	0.091	0.036	0.054
	[-0.085, 0.078]	[-0.107, 0.073]	[-0.095, 0.100]	[-0.003, 0.200]	[-0.090, 0.134]	[-0.066, 0.145]
τ_{BVN}	-0.003	-0.012	-0.004	0.044	0.008	0.012
	[-0.047, 0.049]	[-0.070, 0.044]	[-0.061, 0.052]	[-0.012, 0.105]	[-0.055, 0.068]	[-0.052, 0.066]
$\tau_{MB\text{-}BC,0.05}$	0.015	-0.013	0.000	0.097	0.000	0.007
	[-0.084, 0.118]	[-0.124, 0.116]	[-0.100, 0.122]	[-0.028, 0.220]	[-0.099, 0.138]	[-0.112, 0.144]
$\tau_{MB\text{-}BC,0.25}$	0.015	-0.007	0.012	0.106	0.029	0.034
	[-0.076, 0.113]	[-0.111, 0.108]	[-0.095, 0.116]	[-0.010, 0.226]	[-0.091, 0.154]	[-0.090, 0.142]
P*	0.500	0.500	0.500	0.500	0.500	0.500

Notes: Underweight is defined as BMI below the 20th percentile. For other details, see Table C1.