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

Daniel L. Millimet\textsuperscript{1}  Rusty Tchernis\textsuperscript{2}

\textsuperscript{1}Southern Methodist University & IZA
\textsuperscript{2}Georgia State University & NBER & IZA

April 2012
Researchers in economics and other disciplines are often interested in the causal effect of a binary treatment. In the absence of a random experiment, non-experimental or observational data are used. With such data, non-random selection of subjects into the treatment group must be addressed. The econometric literature on program evaluation is very good at handling the case of selection on the basis of observed attributes. When selection occurs on the basis of unobserved (to the econometrician) attributes, things become murky. Typical strategy is to rely on an instrumental variable. However, a valid instrument is often unavailable. Even if an instrument is available, it may estimate an economically uninteresting parameter (LATE).
Introduction

- This is the situation we confront in this paper

**Research Question:**
What is the causal impact of the participation in the School Breakfast Program (SBP) on childhood obesity?

- As discussed below, we believe participating students differ from non-participants along important and unobserved dimensions
- However, we do not believe we have a strong, credible instrument (as usually conceived) available
Introduction

What do we do in this paper?

1. Compare existing estimators for the estimation of treatment effects when subjects select into treatment on the basis of unobserved attributes, but typical exclusion restrictions are unavailable.

2. Propose two new estimators for use in this context.

3. Assess the causal effect of participation in the SBP on child health
   (a) Using a variety of estimators that do not rely on an exclusion restriction for identification
   (b) Using relatively long-run outcome measures to capture habit formation, behavioral responses
   (c) Using data after late 1990 reforms
Introduction

What do we find on the econometric side?

1. Several estimators are able to circumvent non-random selection despite lack of typical instruments
2. Our new estimators offer advantages in certain situations
3. We offer some practical guidelines to researchers faced with similar situations

What do we find with respect to the School Breakfast Program?

1. We find a positive *association* between SBP and child weight when ignoring non-random selection on unobservables
2. We find a robust, negative *causal effect* of SBP according to all estimators that completely account for non-random selection on unobservables

⇒ SBP is a crucial component in the battle against childhood obesity
Remainder of Talk

1. Brief review of childhood obesity issue
2. Institutional details of school nutrition programs
3. Brief literature review
4. Econometrics of program evaluation
5. Monte Carlo study
6. Application
7. Conclusion
Childhood Obesity

Trends

- Recent media attention has brought obesity, and childhood obesity, into the spotlight
- Obesity rate has increased *threefold* for all children; *fourfold* for 6-11 year-olds since 1970s

**Figure 1. Percentage of U.S. Population That Is Obese**
Childhood Obesity

Trends

- Rise is especially pronounced for low income, minority children

Figure 2. Percentage of Children Who Are Obese and Average BMI among Obese Children, by Group

<table>
<thead>
<tr>
<th>Percent</th>
<th>Average BMI for:</th>
<th>All obese</th>
<th>Low-income obese</th>
<th>African Amer. obese</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>Average BMI for:</td>
<td>27.4</td>
<td>27.1</td>
<td>28.7</td>
</tr>
<tr>
<td></td>
<td>All obese:</td>
<td></td>
<td>Low-income obese:</td>
<td>African Amer. obese:</td>
</tr>
<tr>
<td>20</td>
<td>Average BMI for:</td>
<td>27.0</td>
<td>27.1</td>
<td>28.3</td>
</tr>
<tr>
<td></td>
<td>All obese:</td>
<td></td>
<td>Low-income obese:</td>
<td>African Amer. obese:</td>
</tr>
<tr>
<td>15</td>
<td>Average BMI for:</td>
<td>27.4</td>
<td>27.6</td>
<td>27.8</td>
</tr>
<tr>
<td></td>
<td>All obese:</td>
<td></td>
<td>Low-income obese:</td>
<td>African Amer. obese:</td>
</tr>
<tr>
<td>10</td>
<td>Average BMI for:</td>
<td></td>
<td>28.1</td>
<td>29.1</td>
</tr>
<tr>
<td></td>
<td>All obese:</td>
<td></td>
<td>Low-income obese:</td>
<td>African Amer. obese:</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- All Children
- Low-Income Children
- African American Children

Years

1971–74
1976–80
1988–94
1999–2000
Childhood Obesity

Impacts

- Socio-economic impacts
  - 33% (50%) of overweight preschool- (school-)aged children become obese adults
  - Slew of economic and health impacts of adult obesity

- Medical costs
  - Hospitalization costs due to childhood obesity ≈ $240 million in 2005 (Trasande et al. 2009)
  - Total medical costs due to obesity (adult and child) ≈ $100 billion in 2006

Figure 2: Complications of childhood obesity
Childhood Obesity
Policy Responses

- Policymakers have begun to address childhood obesity
- On February 9, 2010, First Lady Michelle Obama launched the “Let’s Move” campaign to combat the childhood obesity epidemic
- That same day, President Obama signed a memorandum establishing a task force on childhood obesity
- The Presidential memorandum states in part

“Obesity has been recognized as a problem for decades, but efforts to address this crisis to date have been insufficient... [I] have set a goal to solve the problem of childhood obesity within a generation so that children born today will reach adulthood at a healthy weight.”

- Aside from recent policy changes, two important existing programs – particularly for low-income children – include school nutrition programs:
  - School Breakfast Program (SBP)
  - National School Lunch Program (NSLP)
Institutional Details

- For reasons discussed later, our focus is on the SBP, but SBP and NSLP are similarly organized
  - NSLP: developed gradually over time; made permanent by the National School Lunch Act (1946)
  - SBP: established by the Child Nutrition Act (1966); made permanent in 1975 amendments
- Provide nutritionally balanced meals (and afternoon snacks) to children
- Children eligible for free or reduced price meals based on need
- One application to schools covers both programs
- Schools not required to participate in either program (unless state law mandates such)
## Institutional Details

### Eligibility Rules

### Income Guidelines for School Year 2009–2010

<table>
<thead>
<tr>
<th>Household Size</th>
<th>Free Meals</th>
<th>Reduced-Price Meals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maximum Household Income</td>
<td>Maximum Household Income</td>
</tr>
<tr>
<td></td>
<td>(130% of Poverty)</td>
<td>(185% of Poverty)</td>
</tr>
<tr>
<td></td>
<td>Annual</td>
<td>Monthly</td>
</tr>
<tr>
<td>1</td>
<td>$14,079</td>
<td>$1,174</td>
</tr>
<tr>
<td>2</td>
<td>18,941</td>
<td>1,579</td>
</tr>
<tr>
<td>3</td>
<td>23,803</td>
<td>1,984</td>
</tr>
<tr>
<td>4</td>
<td>28,665</td>
<td>2,289</td>
</tr>
<tr>
<td>5</td>
<td>33,527</td>
<td>2,794</td>
</tr>
<tr>
<td>6</td>
<td>38,389</td>
<td>3,200</td>
</tr>
<tr>
<td>7</td>
<td>43,251</td>
<td>3,605</td>
</tr>
<tr>
<td>8</td>
<td>48,113</td>
<td>4,010</td>
</tr>
</tbody>
</table>

Add for each additional person:

- Free Meals:
  - $4,862
  - $406
  - $94
- Reduced-Price Meals:
  - $6,919
  - $577
  - $134
Institutional Details

- Programs are federally funded; organized by the Food & Nutrition Service (FNS) of the USDA
- Administered by state education agencies
- Schools reimbursed per meal served
- Public and private schools may participate
- Federal nutritional requirements established in 1995 School Meals Initiative for Healthy Children (SMI)
  - $\leq 30\%$ of total calories from fat
  - $\leq 10\%$ of total calories from saturated fat
  - SBP: $\geq 25\%$ of RDA for protein, calcium, iron, Vitamins A and C, and calories
  - NSLP: $\geq 33\%$ of RDA for protein, calcium, iron, Vitamins A and C, and calories
- SMI not fully implemented until 1998-1999
### School Meals: Federal Per Meal Reimbursement Rates

*July 1, 2010 - June 30, 2011*

#### School Breakfast Program

<table>
<thead>
<tr>
<th></th>
<th>Non-Severe Need</th>
<th>Severe Need ³</th>
<th>Price of Meals To Children</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Free</strong></td>
<td>$1.48</td>
<td>$1.76</td>
<td>$0</td>
</tr>
<tr>
<td><strong>Reduced Price</strong></td>
<td>$1.18</td>
<td>$1.46</td>
<td>$0.30 (maximum school can charge)</td>
</tr>
<tr>
<td><strong>Paid</strong></td>
<td>$0.26</td>
<td>$0.26</td>
<td>varies ⁴</td>
</tr>
</tbody>
</table>

#### National School Lunch Program ⁵

<table>
<thead>
<tr>
<th></th>
<th>Less than 60%</th>
<th>60% or More ⁶</th>
<th>Price of Meals To Children</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Free</strong></td>
<td>$2.72</td>
<td>$2.74</td>
<td>$0</td>
</tr>
<tr>
<td><strong>Reduced Price</strong></td>
<td>$2.32</td>
<td>$2.34</td>
<td>$0.40 (maximum school can charge)</td>
</tr>
<tr>
<td><strong>Paid</strong></td>
<td>$0.26</td>
<td>$0.28</td>
<td>varies ⁵</td>
</tr>
</tbody>
</table>
SBP is less utilized relative to NSLP.

Details to follow, but roughly speaking...

- 95% of all schools participate in NSLP.
- 86% of those schools also participate in SBP in 2008.
- Participation in SBP is about one-third of NSLP.
Institutional Details
National Participation (SBP): Historical

Figure 1: Student Participation in the Free and Reduced-Price School Breakfast Program

Millimet & Tchernis (SMU & GSU)  Program Evaluation  Apr 2012  16 / 59
Institutional Details
National Participation (NSLP): Historical

Note: F/RP participation only (in millions).
Institutional Details

Underutilization of SBP Relative to NSLP: Historical

![Graph showing the percentage of low-income students in School Breakfast as a percent of low-income students in School Lunch from 1991 to 2005.](image)
Institutional Details

Underutilization of SBP Relative to NSLP: Current
### Institutional Details

National Participation & Costs: SY 2008-2009

<table>
<thead>
<tr>
<th>School Breakfast Program (School Year 2008-2009)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Daily Student Participation</td>
<td>10,846,709</td>
</tr>
<tr>
<td>Free and Reduced-Price Students</td>
<td>8,846,090</td>
</tr>
<tr>
<td>Paid Students</td>
<td>2,000,619</td>
</tr>
<tr>
<td>Change in Free and Reduced-Price Participation in Last 10 Years</td>
<td>59.2%</td>
</tr>
<tr>
<td>Free and Reduced-Price Student Participation Rate (Compared to School Lunch Participation)</td>
<td>46.7</td>
</tr>
<tr>
<td>Rank Among States</td>
<td></td>
</tr>
<tr>
<td>Additional Free and Reduced-Price Students Served if Participation Rate Reached 60%</td>
<td>2,517,386</td>
</tr>
<tr>
<td>Additional Federal Dollars States Would Receive if Participation Rate Reached 60%</td>
<td>$578,898,199</td>
</tr>
<tr>
<td>Number of Schools Participating</td>
<td>86,146</td>
</tr>
<tr>
<td>School Participation Rate (Compared to Number of Schools Serving Lunch)</td>
<td>86.3%</td>
</tr>
<tr>
<td>Federal Funding for School Breakfast</td>
<td>$2,581,964,322</td>
</tr>
<tr>
<td>School Breakfast Mandate in Law (Yes/No)</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>National School Lunch Program (School Year 2008-2009)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Daily Student Participation</td>
<td>30,775,792</td>
</tr>
<tr>
<td>Free and Reduced-Price Students</td>
<td>18,930,484</td>
</tr>
<tr>
<td>Paid Students</td>
<td>11,845,308</td>
</tr>
<tr>
<td>Number of Schools Participating</td>
<td>99,826</td>
</tr>
<tr>
<td>Federal Funding for School Lunch</td>
<td>$8,872,906,763</td>
</tr>
</tbody>
</table>
### Institutional Details
New Proposed Rules: Cost & Nutrition

#### Table 15—Cost (in Millions) of Proposed Rule With Implementation Phase-In Based on LEA Size

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Food Costs</td>
<td>$31.4</td>
<td>$243.3</td>
<td>$443.2</td>
<td>$805.1</td>
<td>$918.4</td>
<td>$2,441.4</td>
</tr>
<tr>
<td>Labor Costs</td>
<td>30.6</td>
<td>237.4</td>
<td>432.5</td>
<td>785.6</td>
<td>896.3</td>
<td>2,382.5</td>
</tr>
<tr>
<td>State Admin</td>
<td>0.1</td>
<td>8.9</td>
<td>9.0</td>
<td>9.3</td>
<td>9.6</td>
<td>36.9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>62.1</td>
<td>489.6</td>
<td>884.8</td>
<td>1,600.0</td>
<td>1,824.4</td>
<td>4,860.9</td>
</tr>
</tbody>
</table>

#### Table 11—Estimated Food Costs by Food Category

[Dollars in millions]

<table>
<thead>
<tr>
<th>Food group</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>$4.4</td>
<td>$29.0</td>
<td>$29.8</td>
<td>$30.5</td>
<td>$31.3</td>
<td>$125.1</td>
</tr>
<tr>
<td>Meat or Meat Alternate</td>
<td>3.1</td>
<td>22.5</td>
<td>24.9</td>
<td>27.6</td>
<td>30.5</td>
<td>108.6</td>
</tr>
<tr>
<td>Fruit Juice</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Fruit (non-juice)</td>
<td>42.3</td>
<td>286.1</td>
<td>301.4</td>
<td>317.1</td>
<td>334.1</td>
<td>1,281.0</td>
</tr>
<tr>
<td>Vegetables</td>
<td>75.6</td>
<td>515.2</td>
<td>547.8</td>
<td>581.2</td>
<td>617.5</td>
<td>2,337.3</td>
</tr>
<tr>
<td>Refined Grains</td>
<td>$116.0</td>
<td>$787.5</td>
<td>$964.7</td>
<td>1,766.5</td>
<td>$1,869.1</td>
<td>$5,503.8</td>
</tr>
<tr>
<td>Whole Grains</td>
<td>91.2</td>
<td>619.3</td>
<td>825.3</td>
<td>1,840.0</td>
<td>1,946.5</td>
<td>5,322.3</td>
</tr>
</tbody>
</table>
Given the rise in childhood obesity, as well as the cost and scope of federal school nutrition programs, a growing literature has emerged assessing these programs.

Identification of causal effects is potentially hindered by:
- Non-random selection of schools into the SBP and NSLP
- Non-random sorting of students across schools
- Non-random selection of students into the SBP and NSLP conditional on attending a participating school
- In combination with a lack of experimental evidence

Three papers have attempted to circumvent this non-random selection in a meaningful way.
Literature Review

- Bhattacharya et al. (JHR 2006)
  - Use NHANES III (early 1990s)
  - Use diff-in-diff (DD) strategy to compare changes between in- and out-of-school periods for children attending schools that do and do not participate in SBP
  - Find no impact on total calories, probability of eating breakfast
  - SBP does improve the nutritional content of breakfast

- Schanzenbach (JHR 2009)
  - Use Early Childhood Longitudinal Study (late 1990s), NHANES III to examine NSLP
  - Two strategies to handle selection: condition on initial weight during fall K, and use RD approach based on income eligibility rules
  - Findings
    - NSLP increases calorie consumption by 40-120 cals/day, probability of being obese by 2%
    - No non-random selection into NSLP conditional on observables
Literature Review

- Millimet et al. (JHR 2010)
  - Use Early Childhood Longitudinal Study
  - Two strategies to address selection based on Altonji et al. (2005)
    - Both strategies entail estimation of treatment effects under the assumption of exogeneity and then assess how much selection on unobservables is required (relative to the amount of selection on observables) to fully explain the estimate
    - First method applies to continuous and binary outcomes; estimates the size of the bias of OLS assuming that the treatment and control group differ in the same way along both observable and unobservable dimensions
    - Second method applies to binary outcomes only; estimates a bivariate probit model constraining the $\rho$ to different values to assess sensitivity of the treatment effect estimate

- Findings
  - Non-random selection by students into the SBP, but not into NSLP
  - Suggestive evidence of a beneficial effect of SBP, but a harmful effect of NSLP, once modest amounts of non-random selection are admitted
Literature Review

- Each of the methods in these papers has its drawbacks

DD approach
- Assumes no time-varying differences across schools by participation status
- Assumes health of students follows the same time trend across schools by participation status
- Defines the treatment as school-level breakfast availability, not student-level participation

RD approach
- Potentially confounds the effects of SBP and NSLP since both have the same eligibility criteria and participation is highly correlated
- With heterogeneous treatment effects, RD estimates the LATE

Altonji et al. approach
- Mostly abandons point estimation in favor of merely suggestive evidence
- Only point estimates identified from bivariate normality assumption
In light of

1. The potential shortcomings of these prior methods,
2. The evidence of non-random selection on unobservables into SBP, but not NSLP,
3. The lack of credible instruments for participation, and
4. The under-utilization of the SBP relative to the NSLP

we focus on the SBP in this paper, and seek an identification strategy that circumvents non-random selection into the treatment group, but does not require a typical exclusion restriction.
Econometric methods for program evaluation using non-experimental data are classified into two groups:

- Selection on observable (SOO) estimation methods
- Selection on unobservable (SOU) estimation methods

Distinction lies in whether there exist unobservable attributes of subjects that are correlated with both treatment assignment and the outcome of interest conditional on the set of observable variables.

If problem is one of SOO, there are many possible estimation strategies (OLS, PS methods, etc.)
If problem is one of SOU, then

1. Use methods that yield a consistent estimate of the ‘effect’ of the treatment *given that one has an exclusion restriction* ... traditional IV

2. Use methods that replace the exclusion restriction with other requirements
   - (a) Covariance Restrictions (Chamberlain, Chamberlain & Griliches 1970s) requires restrictions on the covariance matrix of the system of reduced form eqtns in lieu of typical instruments
   - (b) Higher Moments estimator (Lewbel 1997; Erickson & Whited 2002) requires the endogenous regressor to have an asymmetric dbn, but the structural errors to have a symmetric dbn
   - (c) Identification through Heteroskedasticity (Rigobon 2003; Klein & Vella 2009a,b) requires heteroskedasticity across known regimes or multiplicative heteroskedasticity of unknown form, but depending on the exogenous regressors

3. Obtain a consistent estimate relying on parametric assumptions instead of an exclusion restriction

4. Bound the treatment effect
- Method (1) may be problematic since IVs may only estimate the LATE and IVs are difficult to find in some contexts.

- Method (2) may be problematic as:
  
  (a) Covariance restrictions are difficult to test or justify
  
  (b) Higher Moment IVs may be weak and require large samples
  
  (c) Identification through Heteroskedasticity (KV) appears promising

- Method (3) may be unreliable if parametric assumptions do not hold; IV/CF is approximately linear.

- Method (4) may not be useful in practice (exceptions Lechner 1999; Altonji et al. 2005)
(Specifically) what do we do in this paper?

1. Propose a new method that alters a conventional propensity score-based SOO estimator in such a way so as to minimize the bias when the assumptions underlying SOO fail (i.e., one is really in a SOU world)
   ⇒ Refer to this as the **minimum-biased (MB) estimator**

2. Amend the previous MB estimator by estimating the bias of this estimator and subtracting this from the ‘biased’ estimate
   ⇒ Refer to this as the **bias-corrected minimum-biased (MB-BC) estimator**

3. Compare the performance of these estimators via Monte Carlo with
   - Typical (unadjusted) propensity score-based SOO estimator
   - SOU estimator identified solely from parametric assumptions
   - KV estimator identified through heteroskedasticity

4. Apply our new estimators to study the impact of SBP participation
Setup

- Random sample of $N$ subjects indexed by $i = 1, ..., N$
- $Y_i(T)$ = potential outcome of individual $i$ under treatment $T$, $T \in \{0, 1\}$
- The causal effect: $\tau_i = Y_i(1) - Y_i(0)$
- Parameters of potential interest:
  - Average Treatment Effect (ATE)
  - Average Treatment Effect on the Treated (ATT)
- Formally:
  \[
  \tau_{ATE} = E[\tau_i] = E[Y_i(1) - Y_i(0)]
  \]
  \[
  \tau_{ATT} = E[\tau_i|T = 1] = E[Y_i(1) - Y_i(0)|T = 1]
  \]
  or, conditional on vector of covariates, $X$
  \[
  \tau_{ATE}[X] = E[\tau_i|X] = E[Y_i(1) - Y_i(0)|X]
  \]
  \[
  \tau_{ATT}[X] = E[\tau_i|X, T = 1] = E[Y_i(1) - Y_i(0)|X, T = 1]
  \]
Setup

For each subject, the researcher observes 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 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 \(\tau\) is not trivial.

One assumption is *conditional independence* (CI) or *selection on observables*

\[ Y(0), Y(1) \perp T \mid X, \quad \text{or} \]
\[ Y(0), Y(1) \perp T \mid P(X) \]

where \(P(X_i) = \Pr(T_i = 1 \mid X_i)\) is the propensity score.
Bias When Conditional Independence Fails

- Not surprisingly, estimators that require CI are biased if the CI assumption fails to hold.

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

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

  (A2) Unobservables are trivariate normal: \( \varepsilon_0, \varepsilon_1, u \sim \mathcal{N}_3(0, \Sigma) \), where

  \[
  \Sigma = \begin{bmatrix}
  \sigma_0^2 & \rho_{01} & \rho_{0u} \\
  \rho_{01} & \sigma_1^2 & \rho_{1u} \\
  \rho_{0u} & \rho_{1u} & 1
  \end{bmatrix}
  \]
The bias of the ATT at some value of the propensity score, $P(X)$, is given by

$$B_{ATT}[P(X)] = \hat{\tau}_{ATT}[P(X)] - \tau_{ATT}[P(X)]$$

$$= -\rho_{0u}\sigma_0 \frac{\phi(\Phi^{-1}(P(X)))}{P(X)[1 - P(X)]}$$

where

- $\rho_{0u} = \text{selection on unobservables affecting outcome in untreated state}$ ($\rho_{0u} < 0 \rightarrow \text{positive selection}$)
- $\phi$ and $\Phi$ are standard normal PDF and CDF
- $\hat{\tau}_{ATT}$ is some propensity score based estimator

$B_{ATT}[P(X)]$ is minimized at $P^*(X) = 0.5$
Bias When Conditional Independence Fails

- For the ATE,

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

where

- \( \delta \) = unobserved, individual-specific gain from treatment \((\varepsilon_1 - \varepsilon_0)\)
- \( \rho_{\delta u} \) = selection on unobserved, individual-specific gains \((\rho_{\delta u} < 0 \rightarrow \text{positive selection})\)

- The **bias-minimizing propensity score**, \( P^*(X) \), depends on the error correlation structure

Characterization of $P^*(X)$ and $B_{ATE}[P(X)]$ ...

(i) Two special cases where bias disappears

- If $\rho_{0u}\sigma_0 = 0$, then $\lim_{P(X)\to 1} B_{ATE}[P(X)] = 0$
- If $\rho_{0u}\sigma_0 = -\rho_{\delta u}\sigma_{\delta}$, then $\lim_{P(X)\to 0} B_{ATE}[P(X)] = 0$

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

- above 0.5, but monotonically increasing for $\rho_{\delta u}\sigma_{\delta} > 0$
- 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)

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

- above 0.5, but monotonically decreasing for $\rho_{\delta u}\sigma_{\delta} < 0$
- 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 gains)

$\Rightarrow$ Two types of selection go in same direction, $P^* > 0.5$; when go in opposite direction, $P^* > 0.5$ when selection on unobserved gains is relatively large (absolute value)
Value of $P^*(X)$ Under Different Parameter Values
Estimation

- **Minimum-biased (MB) estimation technique**
  - **Stage 1:** Estimate the propensity score (e.g., probit model)
  - **Stage 2:** Retain only those observations with a propensity score, \( \hat{P}(X_i) \), within a fixed neighborhood around \( P^*(X) \), the bias-minimizing propensity score
  - **Stage 3:** Estimate the ATE or ATT using any propensity-score based estimator that relies on CI using this sub-sample

- **Notes:**
  - Estimator is biased, but it minimizes the bias
  - For ATT... this is straightforward as we know that \( P^*(X) = 0.5 \)
  - For ATE... \( P^*(X) \) is unknown, depends on error correlations
  - If treatment effect is heterogeneous, then interpretation is similar to the LATE; may not be economically interesting
So, for ATE, add **Stage 1.5:** Estimate the error correlations

- Feasible if one places more structure on the model
- Assume functional forms in addition to prior assumptions on separability and normality
- Estimate via OLS

\[
y_i = X_i \beta_0 + X_i T_i (\beta_1 - \beta_0) + \beta_{\lambda_0} (1 - T_i) \left[ \frac{\phi(X_i \gamma)}{1 - \Phi(X_i \gamma)} \right] + \beta_{\lambda_1} T_i \left[ \frac{-\phi(X_i \gamma)}{\Phi(X_i \gamma)} \right] + \eta_i
\]

where \( \phi(\cdot) / \Phi(\cdot) \) is the inverse Mills’ ratio and

\[
\begin{align*}
\beta_{\lambda_0} &= \rho_{0u} \sigma_0 \\
\beta_{\lambda_1} &= \rho_{0u} \sigma_0 + \rho_{\delta u} \sigma_\delta.
\end{align*}
\]

- Replacing \( \gamma \) with \( \hat{\gamma} \) from the first-stage probit yields consistent estimates of \( \rho_{0u} \sigma_0 \) and \( \rho_{\delta u} \sigma_\delta \)
Issues

1. Under the functional form assumptions, OLS estimation of augmented model provides *consistent* estimates of the ATE and ATT even if conditional independence *fails* ... referred to as **BVN estimator**

   ⇒ So, why bother with our estimator? Perhaps our estimator works better when the functional form and/or normality assumptions fail

2. Given estimates of the error correlation structure, the bias of estimators that require conditional independence can be estimated and bias-corrected estimates can be obtained

   ⇒ Example:

   \[
   B_{ATE}[\bar{P}(X)] = - \left\{ \rho_0u\sigma_0 + [1 - \bar{P}^*(X)]\rho_\delta u\sigma_\delta \right\} \left\{ \frac{\phi(\Phi^{-1}(\bar{P}^*(X)))}{P^*(X)[1 - \bar{P}^*(X)]} \right\}
   \]

3. What if normality does not hold? Extend the methods to local deviations from normality using Edgeworth expansions
Bias with Non-normal Errors

Following Lee (1984) ...

- Write the (unknown) joint density of $\varepsilon_j$ and $u$, $f_j(\varepsilon_j, u)$, $j = 0, 1$, as a bivariate Edgeworth series of distributions

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

where $\phi_2$ is the bivariate standard normal density and $A_{rs}$ are functions of the cumulants (or semi-invariants) of $\varepsilon_j$ and $u$

- Consider the case where $r + s \in \{3, 4\}$ and $u \sim \mathbb{N}(0, 1)$
Bias of the ATT is

\[ B_{ATT}[P(X)] = - \left\{ \rho_{0u}\sigma_0 + \kappa_{12}\sigma_0 \frac{h(X)}{2} + \kappa_{13}\sigma_0 \frac{[h(X)^2 - 1]}{6} \right\} \]

\[ \times \left\{ \frac{\phi(h(X))}{\Phi(h(X))[1 - \Phi(h(X))]} \right\} \]

where \( \kappa_{ij} \) are semi-invariants (or cumulants)

Bias of the ATE is

\[ B_{ATE}[P(X)] = - \left\{ \rho_{0u}\sigma_0 + \kappa_{12}\sigma_0 \frac{h(X)}{2} + \kappa_{13}\sigma_0 \frac{[h(X)^2 - 1]}{6} \right\} \]

\[ + [1 - P(X)] \left\{ \rho_{\delta u}\sigma_\delta + \kappa''_{12}\sigma_\delta \frac{h(X)}{2} + \kappa''_{13}\sigma_\delta \frac{[h(X)^2 - 1]}{6} \right\} \]

\[ \times \left\{ \frac{\phi(h(X))}{\Phi(h(X))[1 - \Phi(h(X))]} \right\} \]
Bias of the ATE and ATT is not minimized at a fixed $P^*(X)$

Estimation algorithm requires an altered BVN specification


g_i = X_i\beta_0 + X_i T_i (\beta_1 - \beta_0)

\[+ \beta_{\lambda 01} (1 - T_i) \left( \frac{\phi(X_i \gamma)}{1 - \Phi(X_i \gamma)} \right) + \beta_{\lambda 02} (1 - T_i) \left( \frac{X_i \gamma}{2} \frac{\phi(X_i \gamma)}{1 - \Phi(X_i \gamma)} \right)\]

\[+ \beta_{\lambda 03} (1 - T_i) \left( \frac{(X_i \gamma)^2 - 1}{6} \frac{\phi(X_i \gamma)}{1 - \Phi(X_i \gamma)} \right) + \beta_{\lambda 11} T_i \left( -\frac{\phi(X_i \gamma)}{\Phi(X_i \gamma)} \right)\]

\[+ \beta_{\lambda 12} T_i \left( -\frac{X_i \gamma \phi(X_i \gamma)}{2 \Phi(X_i \gamma)} \right) + \beta_{\lambda 13} T_i \left( \frac{1 - (X_i \gamma)^2}{6} \frac{\phi(X_i \gamma)}{\Phi(X_i \gamma)} \right) + \eta_i\]

where

\[\beta_{\lambda 01} = \rho_{0u} \sigma_0; \quad \beta_{\lambda 11} = \rho_{0u} \sigma_0 + \rho_{\delta u} \sigma_{\delta}\]

\[\beta_{\lambda 02} = \kappa_{12} \sigma_0; \quad \beta_{\lambda 12} = \kappa_{12} \sigma_0 + \kappa'_{12} \sigma_{\delta}\]

\[\beta_{\lambda 03} = \kappa_{13} \sigma_0; \quad \beta_{\lambda 13} = \kappa_{13} \sigma_0 + \kappa''_{13} \sigma_{\delta}\]

Remainder of the estimation algorithm is unchanged
Monte Carlo Experiments

- Goal is to compare MB and MB-BC estimators to
  - BVN estimator relying on functional form assumptions
  - KV estimator relying on heteroskedasticity
    - Parametric TSLS version
    - Heteroskedastic probit used in first-stage $\Rightarrow \hat{P}(X) = \Phi[X\hat{\gamma}/\exp(X\hat{\delta})]$
    - $\hat{P}(X)$ is used as an instrument in the second-stage
  - Typical SOO estimator (Hirano & Imbens (2001) normalized weighting estimators) ... referred to as HI estimator

- The HI estimator is also what we use to operationalize our MB and MB-BC estimators
Setups...

- Two primary experimental designs
  1. Common Effect: $\varepsilon = \varepsilon_0 = \varepsilon_1 \Rightarrow ATE = ATT$
  2. Heterogeneous Effects: $\varepsilon_0 \neq \varepsilon_1 \Rightarrow ATE \neq ATT$

- Within each experimental design, four error structures
  - Trivariate normal
    $\Rightarrow \varepsilon_0, \varepsilon_1, u$ are mean zero, unit var
  - Asymmetric, non-normal multivariate dbn
    $\Rightarrow \varepsilon_0, \varepsilon_1$ are mean zero, unit var, skewness and kurtosis close to a $\chi^2_1$ dbn; $u$ is std normal
  - Repeat above two cases, but now $u$ is heteroskedastic with variance depending on $X$

- At this stage (motivated by our application) we do not consider heterogenous effects due to $X$
Data-generating process...

\[ h(X) = 0.5(x_1 - x_2) + 0.5(x_1^2 - x_2^2) + 2x_1x_2 \]

\[ Y(0) = h(X) + \varepsilon_0 \]

\[ Y(1) = 1 + h(X) + \varepsilon_1 \]

\[ T^* = 0.5 + h(X) - u \]

\[ T = \begin{cases} 
1 & \text{if } T^* > 0 \\
0 & \text{otherwise} 
\end{cases} \]

\[ x_1, x_2 \sim \mathcal{U}(-1, 1) \]
Four specifications for each setup

(S1) Under-specified (linear): $x_1, x_2$
(S2) Correct: $x_1, x_2, x_1^2, x_2^2, x_1 x_2$
(S3) Over-specified: $x_1, x_2, x_1^2, x_2^2, x_1 x_2, x_1^3, x_2^3, x_1 x_2^2, x_1^2 x_2, x_1^2 x_2^2$
(S4) (Really) Over-specified: same as above, but also introduce an irrelevant regressor, $x_3$

250 data sets, $N = 5,000$

Compare root mean squared error (RMSE)
MC findings:

- Under SOO
  - HI estimator does best, but definitely err on the side of over-specifying the model
  - MB is very close and is more robust to under-specifying the model (and, hence, unconfoundedness no longer holding)
    - MB is a useful robustness check under SOO

- Under SOU
  - KV, BVN, MB-BC do best, but definitely err on the side of over-specifying the model (all do very, very poorly in under-specified models)
  - KV tends to do best with heteroskedastic errors; BVN or MB-BC tend to do best otherwise
  - MB-BC outperforms BVN in some cases

- Estimators based on non-normality do not perform as well even under non-normality
- **SOO:**
  - HI best
  - MB not bad

- **SOU:**
  - BVN, MB-BC best with homosked.
  - KV best with heterosked.

- EE always stinks

### Table 1. ATE Estimates in the Common Effect Model ($\tau_i = 1$)

<table>
<thead>
<tr>
<th></th>
<th>Homoskedasticity</th>
<th>Heteroskedasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho_0\sigma_0 - 0$</td>
<td>$\rho_0\sigma_0 - 0.25$</td>
</tr>
<tr>
<td>$\rho_0\sigma_0$ = 1</td>
<td>(1)</td>
<td>(2)</td>
</tr>
</tbody>
</table>

#### I. Normally Distributed Errors

<table>
<thead>
<tr>
<th></th>
<th>Homoskedasticity</th>
<th>Heteroskedasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_{HI}$</td>
<td>0.047</td>
<td>0.440</td>
</tr>
<tr>
<td>$\tau_{MB,0.05}$</td>
<td>0.113</td>
<td>0.407</td>
</tr>
<tr>
<td>$\tau_{MB,0.25}$</td>
<td>0.064</td>
<td>0.402</td>
</tr>
<tr>
<td>$\tau_{MB,EE,0.05}$</td>
<td>0.129</td>
<td>0.471</td>
</tr>
<tr>
<td>$\tau_{MB,EE,0.25}$</td>
<td>0.068</td>
<td>0.437</td>
</tr>
<tr>
<td>$\tau_{KV}$</td>
<td>0.360</td>
<td>0.374</td>
</tr>
<tr>
<td>$\tau_{BVN}$</td>
<td>0.301</td>
<td>0.287</td>
</tr>
<tr>
<td>$\tau_{MB-BC,0.05}$</td>
<td>0.312</td>
<td>0.301</td>
</tr>
<tr>
<td>$\tau_{MB-BC,0.25}$</td>
<td>0.296</td>
<td>0.291</td>
</tr>
<tr>
<td>$\tau_{BVN,EE}$</td>
<td>1.610</td>
<td>1.363</td>
</tr>
<tr>
<td>$\tau_{MB-BC,EE,0.05}$</td>
<td>1.319</td>
<td>1.268</td>
</tr>
<tr>
<td>$\tau_{MB-BC,EE,0.25}$</td>
<td>1.289</td>
<td>1.244</td>
</tr>
</tbody>
</table>

#### II. Asymmetric, Non-Normally Distributed Errors

<table>
<thead>
<tr>
<th></th>
<th>Homoskedasticity</th>
<th>Heteroskedasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_{HI}$</td>
<td>0.050</td>
<td>0.378</td>
</tr>
<tr>
<td>$\tau_{MB,0.05}$</td>
<td>0.097</td>
<td>0.350</td>
</tr>
<tr>
<td>$\tau_{MB,0.25}$</td>
<td>0.059</td>
<td>0.348</td>
</tr>
<tr>
<td>$\tau_{MB,EE,0.05}$</td>
<td>0.130</td>
<td>0.434</td>
</tr>
<tr>
<td>$\tau_{MB,EE,0.25}$</td>
<td>0.069</td>
<td>0.385</td>
</tr>
<tr>
<td>$\tau_{KV}$</td>
<td>0.363</td>
<td>0.400</td>
</tr>
<tr>
<td>$\tau_{BVN}$</td>
<td>0.281</td>
<td>0.308</td>
</tr>
<tr>
<td>$\tau_{MB-BC,0.05}$</td>
<td>0.286</td>
<td>0.313</td>
</tr>
<tr>
<td>$\tau_{MB-BC,0.25}$</td>
<td>0.277</td>
<td>0.303</td>
</tr>
<tr>
<td>$\tau_{BVN,EE}$</td>
<td>1.608</td>
<td>1.631</td>
</tr>
<tr>
<td>$\tau_{MB-BC,EE,0.05}$</td>
<td>1.329</td>
<td>1.514</td>
</tr>
<tr>
<td>$\tau_{MB-BC,EE,0.25}$</td>
<td>1.296</td>
<td>1.503</td>
</tr>
</tbody>
</table>
• SOO:
  ▶ HI best
  ▶ MB not bad

• SOU:
  ▶ BVN, MB-BC best with homosked.
  ▶ KV best with heterosked.

• EE always stinks

---

### Table 2. ATE Estimates in the Heterogeneous Effect Model ($\tau_i = \beta_1 + \delta_i$)

<table>
<thead>
<tr>
<th></th>
<th>Homoskedasticity</th>
<th>Heteroskedasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho_0\sigma_0 \cdot 0$</td>
<td>$\rho_0\sigma_0 \cdot -0.40$</td>
</tr>
<tr>
<td>$\rho_1 \cdot 0.50$</td>
<td>$\rho_1 \cdot -0.50$</td>
<td>$\rho_1 \cdot -0.50$</td>
</tr>
<tr>
<td>$\rho_2\sigma_2 \cdot 0$</td>
<td>$\rho_2\sigma_2 \cdot -0.10$</td>
<td>$\rho_2\sigma_2 \cdot -0.10$</td>
</tr>
</tbody>
</table>

(1) (2) (3) (1) (2) (3)

<table>
<thead>
<tr>
<th>Method</th>
<th>Homoskedasticity</th>
<th>Heteroskedasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_{HI}$</td>
<td>0.046</td>
<td>0.412</td>
</tr>
<tr>
<td>$\tau_{MB,0.05}$</td>
<td>0.100</td>
<td>0.379</td>
</tr>
<tr>
<td>$\tau_{MB,0.25}$</td>
<td>0.062</td>
<td>0.378</td>
</tr>
<tr>
<td>$\tau_{MB,EE,0.05}$</td>
<td>0.124</td>
<td>0.463</td>
</tr>
<tr>
<td>$\tau_{MB,EE,0.25}$</td>
<td>0.066</td>
<td>0.413</td>
</tr>
<tr>
<td>$\tau_{KV}$</td>
<td>0.361</td>
<td>0.369</td>
</tr>
<tr>
<td>$\tau_{BVN}$</td>
<td>0.288</td>
<td>0.262</td>
</tr>
<tr>
<td>$\tau_{MB-BC,0.05}$</td>
<td>0.299</td>
<td>0.271</td>
</tr>
<tr>
<td>$\tau_{MB-BC,0.25}$</td>
<td>0.284</td>
<td>0.259</td>
</tr>
<tr>
<td>$\tau_{BVN,EE}$</td>
<td>1.585</td>
<td>1.708</td>
</tr>
<tr>
<td>$\tau_{MB-BC,EE,0.05}$</td>
<td>1.389</td>
<td>1.431</td>
</tr>
<tr>
<td>$\tau_{MB-BC,EE,0.25}$</td>
<td>1.360</td>
<td>1.401</td>
</tr>
</tbody>
</table>

II. Asymmetric, Non-Normally Distributed Errors

<table>
<thead>
<tr>
<th>Method</th>
<th>Homoskedasticity</th>
<th>Heteroskedasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_{HI}$</td>
<td>0.045</td>
<td>0.380</td>
</tr>
<tr>
<td>$\tau_{MB,0.05}$</td>
<td>0.111</td>
<td>0.304</td>
</tr>
<tr>
<td>$\tau_{MB,0.25}$</td>
<td>0.056</td>
<td>0.311</td>
</tr>
<tr>
<td>$\tau_{MB,EE,0.05}$</td>
<td>0.124</td>
<td>0.496</td>
</tr>
<tr>
<td>$\tau_{MB,EE,0.25}$</td>
<td>0.067</td>
<td>0.397</td>
</tr>
<tr>
<td>$\tau_{KV}$</td>
<td>0.400</td>
<td>0.380</td>
</tr>
<tr>
<td>$\tau_{BVN}$</td>
<td>0.298</td>
<td>0.279</td>
</tr>
<tr>
<td>$\tau_{MB-BC,0.05}$</td>
<td>0.310</td>
<td>0.291</td>
</tr>
<tr>
<td>$\tau_{MB-BC,0.25}$</td>
<td>0.291</td>
<td>0.280</td>
</tr>
<tr>
<td>$\tau_{BVN,EE}$</td>
<td>1.405</td>
<td>1.641</td>
</tr>
<tr>
<td>$\tau_{MB-BC,EE,0.05}$</td>
<td>1.235</td>
<td>1.472</td>
</tr>
<tr>
<td>$\tau_{MB-BC,EE,0.25}$</td>
<td>1.206</td>
<td>1.484</td>
</tr>
</tbody>
</table>
• Not shown here, but worth emphasizing:
  ▶ Under-fitting is very costly; MB is most robust
  ▶ Over-fitting is beneficial for HI, MB, BVN, and MB-BC
  ▶ Over-fitting is somewhat costly for KV

• We use these insights to shape our analysis in the application
Early Childhood Longitudinal Study-K collected by US Dept of Ed
Nationally representative sample of the 1998-1999 kindergarten class
Data is post-SMI reforms
Waves in fall K, spring K, fall 1st, spring 1st, spring 3rd, spring 5th, spring 8th
Retain public school students with non-missing data on age, gender, and SBP participation
• Treatment: SBP participation in spring 1\textsuperscript{st}

• Outcomes:
  ▶ BMI growth rate: fall 1\textsuperscript{st} to spring 3\textsuperscript{rd}, fall 1\textsuperscript{st} to spring 5\textsuperscript{th}
  ▶ Indicators for ‘overweight’ ($>$ 85\textsuperscript{th} percentile) and ‘obese’ ($>$ 95\textsuperscript{th} percentile) in spring 3\textsuperscript{rd} and spring 5\textsuperscript{th}

• Controls (all from fall K wave):
  ▶ Child-specific: race, gender, age (in months), child’s birthweight, NSLP participation
  ▶ Family-specific: region, city type, mother’s age at first birth, WIC benefits prior to K, SES status, mother’s education, maternal employment, food insecurity in household, children’s books at home, imputation dummies

• Summary statistics in Table 3 (in paper)

• $N = 9,952$ of which 3,071 are in the treatment group
Applying estimators discussed above and erring on the side of over-specifying the models yields

1. HI estimator: positive, significant effects on child weight
2. MB estimator: positive, insignificant effects on child weight
3. KV estimator: negative, insignificant effects on child weight
4. BVN estimator: negative, mostly significant effects on child weight
5. MB-BC estimator: negative, mostly significant effects on child weight

Based on our priors, KV, BVN, MB-BC seem to do well in this case; paint a similar picture of the causal effect of participation
### Table 4. Effect of SBP Participation on BMI Growth

<table>
<thead>
<tr>
<th></th>
<th>ATE</th>
<th>ATC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3rd Grade</td>
<td>5th Grade</td>
</tr>
<tr>
<td>( \tau_{OLS} )</td>
<td>0.007</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>[0.003, 0.011]</td>
<td>[0.010, 0.021]</td>
</tr>
<tr>
<td>( \tau_{HI} )</td>
<td>0.010</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>[0.005, 0.014]</td>
<td>[0.015, 0.029]</td>
</tr>
<tr>
<td>( \tau_{MB,0.05} )</td>
<td>0.006</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>[-0.010, 0.016]</td>
<td>[-0.005, 0.034]</td>
</tr>
<tr>
<td>( \tau_{MB,0.25} )</td>
<td>0.007</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>[-0.002, 0.010]</td>
<td>[0.003, 0.021]</td>
</tr>
<tr>
<td>( \tau_{KV} )</td>
<td>-0.029</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td>[-0.051, 0.005]</td>
<td>[-0.057, 0.031]</td>
</tr>
<tr>
<td>( \tau_{BVN} )</td>
<td>-0.032</td>
<td>-0.044</td>
</tr>
<tr>
<td></td>
<td>[-0.057, -0.003]</td>
<td>[-0.082, 0.007]</td>
</tr>
<tr>
<td>( \tau_{MB-BC,0.05} )</td>
<td>-0.031</td>
<td>-0.036</td>
</tr>
<tr>
<td></td>
<td>[-0.063, -0.001]</td>
<td>[-0.088, 0.015]</td>
</tr>
<tr>
<td>( \tau_{MB-BC,0.25} )</td>
<td>-0.031</td>
<td>-0.043</td>
</tr>
<tr>
<td></td>
<td>[-0.061, -0.001]</td>
<td>[-0.082, 0.012]</td>
</tr>
<tr>
<td>( p^* )</td>
<td>0.518</td>
<td>0.612</td>
</tr>
<tr>
<td></td>
<td>[0.310, 0.927]</td>
<td>[0.357, 0.945]</td>
</tr>
</tbody>
</table>
### Table 5. Effect of SBP Participation on Overweight Status

<table>
<thead>
<tr>
<th></th>
<th>ATE</th>
<th></th>
<th>ATT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3rd Grade</td>
<td>5th Grade</td>
<td>3rd Grade</td>
<td>5th Grade</td>
</tr>
<tr>
<td>( \tau_{OLS} )</td>
<td>0.035</td>
<td>0.049</td>
<td>0.035</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>[0.014, 0.054]</td>
<td>[0.032, 0.076]</td>
<td>[0.014, 0.054]</td>
<td>[0.032, 0.076]</td>
</tr>
<tr>
<td>( \tau_{HI} )</td>
<td>0.048</td>
<td>0.076</td>
<td>0.034</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>[0.027, 0.077]</td>
<td>[0.044, 0.106]</td>
<td>[0.009, 0.061]</td>
<td>[0.025, 0.078]</td>
</tr>
<tr>
<td>( \tau_{MB, 0.05} )</td>
<td>0.003</td>
<td>0.057</td>
<td>-0.002</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>[-0.058, 0.094]</td>
<td>[-0.053, 0.116]</td>
<td>[-0.044, 0.097]</td>
<td>[-0.053, 0.110]</td>
</tr>
<tr>
<td>( \tau_{MB, 0.25} )</td>
<td>0.016</td>
<td>0.030</td>
<td>0.018</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>[-0.013, 0.049]</td>
<td>[-0.003, 0.065]</td>
<td>[-0.018, 0.043]</td>
<td>[0.007, 0.077]</td>
</tr>
<tr>
<td>( \tau_{KV} )</td>
<td>-0.136</td>
<td>-0.206</td>
<td>-0.136</td>
<td>-0.206</td>
</tr>
<tr>
<td></td>
<td>[-0.283, 0.059]</td>
<td>[-0.359, 0.038]</td>
<td>[-0.283, 0.059]</td>
<td>[-0.359, 0.038]</td>
</tr>
<tr>
<td>( \tau_{BVN} )</td>
<td>-0.208</td>
<td>-0.278</td>
<td>-0.088</td>
<td>-0.108</td>
</tr>
<tr>
<td></td>
<td>[-0.329, -0.044]</td>
<td>[-0.397, -0.095]</td>
<td>[-0.163, 0.007]</td>
<td>[-0.182, 0.004]</td>
</tr>
<tr>
<td>( \tau_{MB-BC, 0.05} )</td>
<td>-0.226</td>
<td>-0.255</td>
<td>-0.230</td>
<td>-0.296</td>
</tr>
<tr>
<td></td>
<td>[-0.361, -0.019]</td>
<td>[-0.435, -0.060]</td>
<td>[-0.357, -0.021]</td>
<td>[-0.453, -0.027]</td>
</tr>
<tr>
<td>( \tau_{MB-BC, 0.25} )</td>
<td>-0.213</td>
<td>-0.282</td>
<td>-0.211</td>
<td>-0.263</td>
</tr>
<tr>
<td></td>
<td>[-0.332, -0.031]</td>
<td>[-0.397, -0.086]</td>
<td>[-0.361, -0.021]</td>
<td>[-0.413, -0.053]</td>
</tr>
<tr>
<td>( P^* )</td>
<td>0.502</td>
<td>0.534</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td></td>
<td>[0.314, 0.835]</td>
<td>[0.365, 0.818]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_{OLS}$</td>
<td>$\tau_{HI}$</td>
<td>$\tau_{MB,0.05}$</td>
<td>$\tau_{MB,0.25}$</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td>0.028 [0.008, 0.043]</td>
<td>0.028 [0.008, 0.049]</td>
<td>0.001 [-0.063, 0.080]</td>
<td>0.015 [-0.015, 0.043]</td>
</tr>
<tr>
<td></td>
<td>0.039 [0.024, 0.062]</td>
<td>0.053 [0.026, 0.080]</td>
<td>0.068 [-0.031, 0.103]</td>
<td>0.041 [-0.005, 0.063]</td>
</tr>
<tr>
<td></td>
<td>0.028 [0.008, 0.043]</td>
<td>0.032 [0.008, 0.052]</td>
<td>0.044 [-0.037, 0.105]</td>
<td>0.023 [-0.003, 0.048]</td>
</tr>
<tr>
<td></td>
<td>0.039 [0.024, 0.062]</td>
<td>0.044 [0.022, 0.070]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Effect of SBP Participation on Obesity Status
Conclusion

- Confronted with the desire to use non-experimental data to evaluate the causal effect of SBP participation on child health where children self-select into the program on the basis of unobservables but we lack a valid exclusion restriction
- This is a frequent occurrence in applied econometrics
- Here, we propose two new methods for applicable in this situation
  - MB minimizes the bias, MB-BC is a bias-corrected version
  - Compare the performance of these new estimators to several existing estimators: HI, KV, and BVN
- Results
  - MB, MB-BC appear to be nice complements to existing estimators
  - KV, BVN, and MB-BC all suggest a negative and marginally statistically significant causal effect of SBP on child weight
Potential future work on the econometric side:

1. Differential weighting around $P^*(X)$ using (non-uniform) kernels
2. Assessment when $\tau_i$ varies with $X_i$
3. Develop estimator to minimize the bias of the $ATE$ or $ATE$, rather than $ATE(X)$ or $ATT(X)$ ... think about BC estimate at each $P$ and then integrating over $P$ to get a BC estimate of the ‘unconditional’ $ATE$ (or $ATT$).

Potential future work on the evaluation of nutrition programs:

1. Consider the effects of multiple program participation: SBP, NSLP, and SNAP
2. Account for measurement error in participation
3. Consider other outcomes: cognitive, non-cognitive measures