The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$115	1-year subscription	$145
2-year subscription	$210	2-year subscription	$270
3-year subscription	$285	3-year subscription	$375
1-year student subscription	$ 85	1-year student subscription	$115
1-year institutional subscription	$345	1-year institutional subscription	$375
2-year institutional subscription	$625	2-year institutional subscription	$685
3-year institutional subscription	$875	3-year institutional subscription	$965
Electronic only		**Electronic only**	
1-year subscription	$ 85	1-year subscription	$ 85
2-year subscription	$155	2-year subscription	$155
3-year subscription	$215	3-year subscription	$215
1-year student subscription	$ 55	1-year student subscription	$ 55

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, STATA, Stata Press, Mata, MATA, and NetCourse are registered trademarks of StataCorp LP.

The Stata Journal (2015)
15, Number 2, pp. 411–436

Bounding treatment effects: A command for the partial identification of the average treatment effect with endogenous and misreported treatment assignment

Ian McCarthy
Emory University
Atlanta, GA
ian.mccarthy@emory.edu

Daniel L. Millimet
Southern Methodist University
Dallas, TX
and IZA
Bonn, Germany
millimet@smu.edu

Manan Roy
University of North Carolina
Chapel Hill, NC
mananroy@email.unc.edu

Abstract. We present a new command, `tebounds`, that implements a variety of techniques to bound the average treatment effect of a binary treatment on a binary outcome in light of endogenous and misreported treatment assignment. To tighten the worst case bounds, the monotone treatment selection, monotone treatment response, and monotone instrumental-variable assumptions of Manski and Pepper (2000, *Econometrica* 68: 997–1010), Kreider and Pepper (2007, *Journal of the American Statistical Association* 102: 432–441), Kreider et al. (2012, *Journal of the American Statistical Association* 107: 958–975), and Gundersen, Kreider, and Pepper (2012, *Journal of Econometrics* 166: 79–91) may be imposed. Imbens–Manski confidence intervals are provided.

Keywords: st0386, tebounds, treatment effects, selection, misreporting, monotone instrumental variable, monotone treatment selection, monotone treatment response, partial identification, set identification

1 Introduction

The causal effects of binary treatment on an outcome of interest are a central component of empirical research in economics and many other disciplines. When individual units self select into treatment and when prospective randomization of the treatment and control groups is not feasible, researchers must rely on observational data and adopt alternative empirical methods intended to control for the inherent self selection.

If individual units self select on the basis of observed variables (selection on observed variables), a variety of appropriate methodologies exist to estimate the causal effects of the treatment. If, instead, individuals self select on the basis of unobserved variables (selection on unobserved variables), estimation of causal effects is more difficult. In such cases, strong assumptions are often needed to achieve point identification. The credibility of such approaches is further diminished if treatment assignment is misreported. Moreover, as shown in Millimet (2011), estimators appropriate for situations characterized by selection on observed variables may perform poorly when treatment assignment is misreported.

An alternative approach is to abandon the goal of point identification and instead seek to partially identify (or set identify) causal effects.[1] Partial identification approaches are more often credible than those yielding point identification because they highlight what may be learned without invoking perhaps untenable assumptions. As Manski (2013, 2–3) states: "Exact predictions are common, and expressions of uncertainty are rare. Yet policy predictions often are fragile. Conclusions may rest on critical unsupported assumptions or on leaps of logic. Then the certitude of policy analysis is not credible." Similarly, Bontemps, Magnac, and Maurin (2012, 1129) write: "Point identification is often achieved by using strong and difficult to motivate restrictions on the parameters of interest." Tamer (2010, 168) states: "Stronger assumptions will lead to more information about a parameter, but less credible inferences can be conducted." Instead, Manski (2013, 3) advocates the "honest portrayal of partial knowledge".

While it is certainly true that partial identification approaches may be less satisfying than estimators that yield point identification, there is much that can be learned without imposing stringent assumptions. Tamer (2010, 168) argues that "models that do not point identify parameters of interest can, and typically do, contain valuable information about these parameters". In particular, bounds on the parameter of interest often exclude 0, thereby identifying the sign of the parameter. Moreover, bounds can exclude extreme values that may, for example, be useful in determining whether a program fails a cost–benefit analysis.

Here we provide a means to partially identify the average treatment effect (ATE) of a binary treatment on a binary outcome under a variety of assumptions concerning the nature of the self-selection process and the nature and frequency of misreporting of treatment assignment.

2 Framework and methodology

Focusing on binary outcomes, the ATE is given by

$$\text{ATE}(1,0) = P\{Y(1)=1|X \in \Omega\} - P\{Y(0)=1|X \in \Omega\} \tag{1}$$

where $Y(1)$ denotes the outcome if an individual unit receives the treatment, denoted by $D^* = 1$, and $Y(0)$ denotes the outcome if an individual unit does not receive the treatment, denoted by $D^* = 0$. $Y(1)$ and $Y(0)$ are potential outcomes because only one is realized for any given individual. The observed outcome for a particular individual is given by $Y = D^*Y(1) + (1 - D^*)Y(0)$. X denotes a vector of observed covariates

1. Bontemps, Magnac, and Maurin (2012, 1129) state: "A parameter is set identified when the identifying restrictions impose that it lies in a set that is smaller than its potential domain."

whose values lie in the set Ω. To simplify notation, the conditioning on X is left implicit throughout the remainder of the article.

The terms in (1) can be written as

$$P\{Y(1)=1\} = P\{Y(1)=1|D^*=1\}P(D^*=1) + P\{Y(1)=1|D^*=0\}P(D^*=0) \tag{2}$$

$$P\{Y(0)=1\} = P\{Y(0)=1|D^*=1\}P(D^*=1) + P\{Y(0)=1|D^*=0\}P(D^*=0) \tag{3}$$

Two problems arise in the identification of the ATE. The first is referred to as the selection problem. If receipt of treatment is observed, then the sampling process itself identifies the selection probability $P(D^*=1)$, the censoring probability $P(D^*=0)$, and the expectation of outcomes conditional on the outcome being observed, $P\{Y(1)=1|D^*=1\}$ and $P\{Y(0)=1|D^*=0\}$. But the sampling process cannot identify the counterfactual probabilities, $P\{Y(1)=1|D^*=0\}$ and $P\{Y(0)=1|D^*=1\}$, and so $P\{Y(1)=1\}$ and $P\{Y(0)=1\}$ are not point identified by the sampling process alone. The second is referred to as the problem of measurement or classification error. True treatment status may not be observed for all observations. Instead of observing D^*, the indicator D is observed. If $D \neq D^*$ for all units, the sampling process alone does not provide any useful information on true treatment status D^*, and all the probabilities on the right-hand side of (2) and (3) are unknown.

To proceed, we present some notation and preliminaries. First, note the following identities:

$$P\{Y(1)=1|D^*=1\} = P(Y=1|D^*=1)$$
$$P\{Y(1)=0|D^*=1\} = P(Y=0|D^*=1)$$
$$P\{Y(0)=1|D^*=0\} = P(Y=1|D^*=0)$$
$$P\{Y(0)=0|D^*=0\} = P(Y=0|D^*=0)$$

Second, let the latent variable Z^* denote whether reported treatment assignment is accurate, where $Z^*=1$ if $D^*=D$ and $Z^*=0$ otherwise. Third, define the following notation:

$\theta_1^+ \equiv P(Y=1, D=1, Z^*=0)$

$\Rightarrow$ fraction of observations that are false positives with $Y=1$

$\theta_0^+ \equiv P(Y=0, D=1, Z^*=0)$

$\Rightarrow$ fraction of observations that are false positives with $Y=0$

$\theta_1^- \equiv P(Y=1, D=0, Z^*=0)$

$\Rightarrow$ fraction of observations that are false negatives with $Y=1$

$\theta_0^- \equiv P(Y=0, D=0, Z^*=0)$

$\Rightarrow$ fraction of observations that are false negatives with $Y=0$

We can now further decompose $P\{Y(1)=1\}$ and $P\{Y(0)=1\}$ as follows:

$$\begin{aligned}
P\{Y(1)=1\} &= \frac{P(Y=1,D^*=1)}{P(D^*=1)}P(D^*=1)+P\{Y(1)=1|D^*=0\}P(D^*=0)\\
&= \{P(Y=1,D=1)-\theta_1^+ +\theta_1^-\}\\
&\quad + P\{Y(1)=1|D^*=0\}\{P(D=0)+(\theta_1^+ +\theta_0^+)-(\theta_1^- +\theta_0^-)\} \qquad (4)\\
P\{Y(0)=1\} &= P\{Y(0)=1|D^*=1\}P(D^*=1)+\frac{P(Y=1,D^*=0)}{P(D^*=0)}P(D^*=0)\\
&= P\{Y(0)=1|D^*=1\}\{P(D=1)-(\theta_1^+ +\theta_0^+)+(\theta_1^- +\theta_0^-)\}\\
&\quad + \{P(Y=1,D=0)+\theta_1^+ -\theta_1^-\} \qquad (5)
\end{aligned}$$

Unless assumptions are imposed on the nature of the selection problem, the missing counterfactual terms ($P\{Y(1)=1|D^*=0\}$ and $P\{Y(0)=1|D^*=1\}$) are not identified. Similarly, absent further assumptions, the terms representing the extent of measurement error ($\theta_1^+,\theta_0^+,\theta_1^-,\theta_0^-$) are not identified.

2.1 Assumptions

We now turn to the assumptions imposed to bound the ATE in the presence of selection and measurement error. For additional discussion regarding these assumptions and their practical implications, see Manski and Pepper (2000), Kreider and Pepper (2007), Kreider et al. (2012), and Gundersen, Kreider, and Pepper (2012). We stress that the appropriateness of these assumptions will tend to vary by application. For example, in a study of food stamp programs, respondents may be unlikely to report receiving food stamps when in fact they do not, in which case an assumption of no false positives may be appropriate. Users of the `tebounds` command should critically consider these various assumptions when interpreting their results.

Assumptions regarding measurement error include the following:

A1. Arbitrary errors with the upper bound (UB), $P(Z^*=0)\le Q$. Under this assumption, we have the following general conditions on the measurement error parameters:

$$\begin{gathered}
0\le\theta_1^-\le\min\{Q,P(Y=1,D=0)\}\equiv\theta_1^{\mathrm{UB}-}\\
0\le\theta_0^-\le\min\{Q,P(Y=0,D=0)\}\equiv\theta_0^{\mathrm{UB}-}\\
0\le\theta_1^+\le\min\{Q,P(Y=1,D=1)\}\equiv\theta_1^{\mathrm{UB}+}\\
0\le\theta_0^+\le\min\{Q,P(Y=0,D=1)\}\equiv\theta_0^{\mathrm{UB}+}\\
\theta_1^+ +\theta_1^- +\theta_0^+ +\theta_0^-\le Q
\end{gathered}$$

A2. No false positives, $P(Z^*=1|D=1)=1$. This assumption simplifies (4) and (5) above because $\theta_1^+=\theta_0^+=0$.

Assumptions regarding the selection process include the following:

a1. exogenous selection

a2. worst-case selection (no assumption about selection)

a3. monotone treatment selection (MTS)

a4. MTS and monotone treatment response (MTR)

a5. monotone instrumental variable (MIV) and MTS

a6. MIV + MTS + MTR

We discuss each of these assumptions about the selection process in more detail throughout the remainder of this section.

2.2 Exogenous selection bounds

The assumption of exogenous selection implies that

$$P\{Y(1)=1\} = P\{Y(1)=1|D^*=1\} = P\{Y(1)=1|D^*=0\}$$
$$P\{Y(0)=1\} = P\{Y(0)=1|D^*=1\} = P\{Y(0)=1|D^*=0\}$$

In this case, (2) and (3) become, respectively,

$$\begin{aligned} P\left\{Y(1)=1\right\} &= P\left\{Y=1|D^*=1\right\}P(D^*=1) + P\left\{Y(1)=1|D^*=0\right\}P(D^*=0) \\ &= P\left\{Y=1|D^*=1\right\}P(D^*=1) + P\left\{Y=1|D^*=1\right\}\left\{1-P(D^*=1)\right\} \\ &= P\left\{Y=1|D^*=1\right\}\left\{P(D^*=1)-P(D^*=1)+1\right\} \\ &= P\left(Y=1|D^*=1\right) \end{aligned}$$

and

$$\begin{aligned} P\left\{Y(0)=1\right\} &= P\left\{Y(0)=1|D^*=1\right\}P(D^*=1) + P\left(Y=1|D^*=0\right)P(D^*=0) \\ &= P\left(Y=1|D^*=0\right)\left\{1-P(D^*=0)\right\} + P\left(Y=1|D^*=0\right)P(D^*=0) \\ &= P\left(Y=1|D^*=0\right)\left\{P(D^*=0)-P(D^*=0)+1\right\} \\ &= P\left(Y=1|D^*=0\right) \end{aligned}$$

The ATE is then given by $P(Y=1|D^*=1) - P(Y=1|D^*=0)$, which is point identified in the absence of measurement error. However, allowing for measurement error, D^* is unobserved and these quantities can be written as

$$\begin{aligned} P(Y=1|D^*=1) &= \frac{P(Y=1, D^*=1)}{P(D^*=1)} \\ &= \frac{P(Y=1, D=1) - \theta_1^+ + \theta_1^-}{P(D=1) - (\theta_1^+ + \theta_0^+) + (\theta_1^- + \theta_0^-)} \end{aligned} \tag{6}$$

$$\begin{aligned} P(Y=1|D^*=0) &= \frac{P(Y=1, D^*=0)}{P(D^*=0)} \\ &= \frac{P(Y=1, D=0) + \theta_1^+ - \theta_1^-}{P(D=0) + (\theta_1^+ + \theta_0^+) - (\theta_1^- + \theta_0^-)} \end{aligned} \tag{7}$$

Thus, point identification is not possible except under strict assumptions about the values of the unknown quantities in (6) and (7).

Instead, we may use assumptions A1 and A2 to get the lower bound (LB) and UB on the ATE. For example, following proposition 1 in Kreider and Pepper (2007), we know that

$$\begin{aligned} P(Y=1|D^*=1) &\in \left\{ \frac{P(Y=1, D=1) - \delta}{P(D=1) - 2\delta + Q}, \frac{P(Y=1, D=1) + \gamma}{P(D=1) + 2\gamma - Q} \right\} \\ &\equiv (\mathrm{LB}_1, \mathrm{UB}_1) \end{aligned}$$

where

$$\delta = \begin{cases} \theta_1^{\mathrm{UB}+} & \text{if } P(Y=1, D=1) - P(Y=0, D=1) \\ & \quad -Q \quad \leq 0 \\ \max\{0, Q - P(Y=0, D=0)\} & \text{otherwise} \end{cases}$$

$$\gamma = \begin{cases} \theta_1^{\mathrm{UB}-} & \text{if } P(Y=1, D=1) - P(Y=0, D=1) \\ & \quad +Q \quad \leq 0 \\ \max\{0, Q - P(Y=0, D=1)\} & \text{otherwise} \end{cases}$$

and

$$\begin{aligned} P(Y=1|D^*=0) &\in \left\{ \frac{P(Y=1, D=0) - \delta}{P(D=0) - 2\delta + Q}, \frac{P(Y=1, D=0) + \gamma}{P(D=0) + 2\gamma - Q} \right\} \\ &\equiv (\mathrm{LB}_0, \mathrm{UB}_0) \end{aligned}$$

where

$$\delta = \begin{cases} \theta_1^{\mathrm{UB}-} & \text{if } P(Y=1, D=0) - P(Y=0, D=0) \\ & \quad -Q \quad \leq 0 \\ \max\{0, Q - P(Y=0, D=1)\} & \text{otherwise} \end{cases}$$

$$\gamma = \begin{cases} \theta_1^{\mathrm{UB}+} & \text{if } P(Y=1, D=0) - P(Y=0, D=0) \\ & \quad +Q \quad \leq 0 \\ \max\{0, Q - P(Y=0, D=0)\} & \text{otherwise} \end{cases}$$

Bounds on the ATE are then given by

$$\text{ATE} \in (\text{LB}_1 - \text{UB}_0, \text{UB}_1 - \text{LB}_0)$$

However, these bounds are not sharp because it is possible that a different set of measurement error parameters will maximize (minimize) the difference to get the UB (LB) on the ATE. Also, $\theta_1^+ + \theta_1^- + \theta_0^+ + \theta_0^- \leq Q$ has yet to be imposed. Accounting for both of these issues, Proposition A.1 in Kreider and Pepper (2007) shows that

$$\inf_{b\in\left(0,\theta_1^{\text{UB}+}\right),\widetilde{b}\in\left\{0,\min(Q-b,\theta_0^{\text{UB}-})\right\}} \left\{ \frac{P\left(Y=1,D=1\right)-b}{P(D=1)-b+\widetilde{b}} - \frac{P\left(Y=1,D=0\right)+b}{P(D=0)+b-\widetilde{b}} \right\} \tag{8}$$

$$\leq \text{ATE}$$

$$\leq \sup_{a\in(0,\theta_1^{\text{UB}-}),\widetilde{a}\in\left\{0,\min(Q-a,\theta_0^{\text{UB}+})\right\}} \left\{ \frac{P\left(Y=1,D=1\right)+a}{P(D=1)+a-\widetilde{a}} - \frac{P\left(Y=1,D=0\right)-a}{P(D=0)-a+\widetilde{a}} \right\}$$

Estimation follows by performing separate two-way grid searches for $(b, \widetilde{b})$ and $(a, \widetilde{a})$ over the feasible region, where[2]

$$\begin{aligned} b &\in [0, \min\{Q, P\left(Y=1, D=1\right)\}] \\ \widetilde{b} &\in [0, \min\{Q-b, P\left(Y=0, D=0\right)\}] \\ a &\in [0, \min\{Q, P\left(Y=1, D=0\right)\}] \\ \widetilde{a} &\in [0, \min\{Q-a, P\left(Y=0, D=1\right)\}] \end{aligned}$$

In our `tebounds` command, the granularity of the grid search is dictated by the `np()` option. Note that the results of the grid search for $(b, \widetilde{b})$ may not separately minimize LB_1 and maximize UB_0; however, by focusing on the difference $(\text{LB}_1 - \text{UB}_0)$, we ensure that the same $(b, \widetilde{b})$ are used in LB_1 and UB_0, which ultimately provides tighter bounds than if we allowed $(b, \widetilde{b})$ to vary across LB_1 and UB_0, and similarly for $(a, \widetilde{a})$.

If we further impose the assumption of no false positives, (6) and (7) simplify to

$$\begin{aligned} P\left(Y=1|D^*=1\right) &= \frac{P(Y=1, D=1) + \theta_1^-}{P(D=1) + (\theta_1^- + \theta_0^-)} \\ P\left(Y=1|D^*=0\right) &= \frac{P(Y=1, D=0) - \theta_1^-}{P(D=0) - (\theta_1^- + \theta_0^-)} \end{aligned}$$

2. The two-way grid search allows the bounds for $\widetilde{b}$ and $\widetilde{a}$ to vary based on the proposed values of b and a, respectively, as well as Q. Without the inclusion of $\widetilde{b}$ and $\widetilde{a}$, the bounds as presented in Kreider and Pepper (2007) do not necessarily satisfy $\theta_1^+ + \theta_1^- + \theta_0^+ + \theta_0^- \leq Q$ (Kreider, Pepper, and Roy 2013). We are grateful to Brent Kreider for his comments and assistance.

Bounds on $P\left(Y=1|D^{*}=1\right)$ are then given by

$$P\left(Y=1|D^{*}=1\right)\in\left\{\frac{P(Y=1,D=1)}{P(D=1)+\theta_0^{\text{UB}-}},\frac{P(Y=1,D=1)+\theta_1^{\text{UB}-}}{P(D=1)+\theta_1^{\text{UB}-}}\right\}$$
$$\equiv\left(\text{LB}_1,\text{UB}_1\right)$$

Similarly, bounds on $P\left(Y=1|D^{*}=0\right)$ are given by

$$P\left(Y=1|D^{*}=0\right)\in\left\{\frac{P\left(Y=1,D=0\right)-\theta_1^{\text{UB}-}}{P(D=0)-\theta_1^{\text{UB}-}},\frac{P\left(Y=1,D=0\right)}{P(D=0)-\theta_0^{\text{UB}-}}\right\}$$
$$\equiv\left(\text{LB}_0,\text{UB}_0\right)$$

Bounds on the ATE are again given by ATE $\in\left(\text{LB}_1-\text{UB}_0,\text{UB}_1-\text{LB}_0\right)$, which yields

$$\begin{aligned}&\inf_{h\in\left(0,\theta_0^{\text{UB}-}\right)}\left\{\frac{P\left(Y=1,D=1\right)}{P(D=1)+h}-\frac{P\left(Y=1,D=0\right)}{P(D=0)-h}\right\}\\&\leq\text{ATE}\\&\leq\sup_{a\in\left(0,\theta_1^{\text{UB}-}\right)}\left\{\frac{P\left(Y=1,D=1\right)+a}{P(D=1)+a}-\frac{P\left(Y=1,D=0\right)-a}{P(D=0)-a}\right\}\end{aligned}\tag{9}$$

Estimation follows by performing separate grid searches for h and a over the feasible region, ensuring that LB_0, LB_1, UB_0, and UB_1 do not exceed 1.

2.3 Worst-case selection bounds

The worst-case bounds are obtained without invoking any assumptions; only the sampling process is used. The ATE is given by (1), and the components of the ATE are in (2) and (3). Using the fact that the missing counterfactuals in (2) and (3) are bounded by 0 and 1, we know that

$$P\left\{Y(1)=1\right\}\in\left\{P\left(Y=1,D^{*}=1\right),P\left(Y=1,D^{*}=1\right)+P(D^{*}=0)\right\}$$
$$P\left\{Y(0)=1\right\}\in\left\{P\left(Y=1,D^{*}=0\right),P(D^{*}=1)+P\left(Y=1,D^{*}=0\right)\right\}$$

in the absence of measurement error, and bounds on the ATE follow as above.

Allowing for measurement error, the bounds become

$$\begin{aligned}P\{Y(1)=1\}\in\big\{&P(Y=1,D=1)-\theta_1^{+}+\theta_1^{-},\\&P(Y=1,D=1)+P(D=0)+\theta_0^{+}-\theta_0^{-}\big\}\\P\big\{Y(0)=1\big\}\in\big\{&P(Y=1,D=0)+\theta_1^{+}-\theta_1^{-},\\&P(Y=1,D=0)+P(D=1)-\theta_0^{+}+\theta_0^{-}\big\}\end{aligned}$$

In this case, under the assumption of arbitrary errors,

$$\begin{aligned} P\{Y(1)=1\} &\in \{P(Y=1,D=1)-\theta_1^{\text{UB}+}, P(Y=1,D=1)+P(D=0)+\theta_0^{\text{UB}+}\} \\ &\equiv (\text{LB}_1, \text{UB}_1) \\ P\{Y(0)=1\} &\in \{P(Y=1,D=0)-\theta_1^{\text{UB}-}, P(Y=1,D=0)+P(D=1)+\theta_0^{\text{UB}-}\} \\ &\equiv (\text{LB}_0, \text{UB}_0) \end{aligned}$$

which yields the following bounds for the ATE:

$$\begin{aligned} \text{ATE} \in \; & \{P(Y=1,D=1)-\min(Q,\theta_1^{\text{UB}+}+\theta_0^{\text{UB}-}) \\ & \quad - P(Y=1,D=0)-P(D=1), \\ & P(Y=1,D=1)+P(D=0)+\min(Q,\theta_0^{\text{UB}+}+\theta_1^{\text{UB}-}) \\ & \quad - P(Y=1,D=0)\} \end{aligned} \tag{10}$$

Under the assumption of no false positives,

$$\begin{aligned} P\{Y(1)=1\} &\in \{P(Y=1,D=1), P(Y=1,D=1)+P(D=0)\} \\ &\equiv (\text{LB}_1, \text{UB}_1) \\ P\{Y(0)=1\} &\in \{P(Y=1,D=0)-\theta_1^{\text{UB}-}, P(Y=1,D=0)+P(D=1)+\theta_0^{\text{UB}-}\} \\ &\equiv (\text{LB}_0, \text{UB}_0) \end{aligned}$$

and the previous bounds simplify to

$$\begin{aligned} \text{ATE} \in \; & \{P(Y=1,D=1)-P(Y=1,D=0)-P(D=1)-\theta_0^{\text{UB}-}, \\ & P(Y=1,D=1)+P(D=0)-P(Y=1,D=0)+\theta_1^{\text{UB}-}\} \end{aligned} \tag{11}$$

2.4 MTS

The worst-case bounds may be tightened if we are willing to impose some assumptions on the nature of the selection process. The MTS assumption assumes that expected potential outcomes move in a particular direction when comparing individuals in the treatment and control groups. We consider two cases: negative and positive selection. However, we must be cautious in interpreting each case depending on whether the outcome, Y, is desirable. Here we assume that $Y=1$ corresponds to the desirable outcome.

Negative monotone treatment selection (MTSn)

MTSn refers to the case of negative selection. In this case, individuals in the treatment group are more likely to experience a bad outcome conditional on treatment assignment. Being that $Y=1$ denotes a good outcome, MTSn implies that

$$\begin{aligned} P\{Y(1)=1|D^*=0\} &\geq P\{Y(1)=1|D^*=1\} \\ P\{Y(0)=1|D^*=0\} &\geq P\{Y(0)=1|D^*=1\} \end{aligned}$$

We can then rewrite $P\{Y(1)=1\}$ as

$$\begin{aligned}P\{Y(1)=1\} &= P\{Y(1)=1|D^*=1\}P(D^*=1)+P\{Y(1)=1|D^*=0\}P(D^*=0)\\ &= P(Y=1|D^*=1)\{1-P(D^*=0)\}+P\{Y(1)=1|D^*=0\}\\ &\quad P(D^*=0)\\ &= P(Y=1|D^*=1)+P(D^*=0)[P\{Y(1)=1|D^*=0\}\\ &\quad -P(Y=1|D^*=1)]\end{aligned}$$

where the final term is nonnegative under MTSn. This implies that $P\{Y(1)=1\} \geq P(Y=1|D^*=1)$. The LB for $P\{Y(1)=1\}$ is therefore obtained assuming $P\{Y(1)=1|D^*=0\}=P(Y=1|D^*=1)$, and the UB is obtained assuming $P\{Y(1)=1|D^*=0\}=1$. This yields

$$P\{Y(1)=1\} \in \left\{\frac{P(Y=1,D^*=1)}{P(D^*=1)}, P(D^*=0)+P(Y=1,D^*=1)\right\}$$

A similar inspection of $P\{Y(0)=1\}$ yields

$$\begin{aligned}P\{Y(0)=1\} &= P\{Y(0)=1|D^*=1\}P(D^*=1)+P\{Y(0)=1|D^*=0\}P(D^*=0)\\ &= P\{Y(0)=1|D^*=1\}P(D^*=1)+P(Y=1|D^*=0)\\ &\quad \{1-P(D^*=1)\}\\ &= P\{Y=1|D^*=0\}+P(D^*=1)[P\{Y(0)=1|D^*=1\}\\ &\quad -P\{Y(0)=1|D^*=0\}]\end{aligned}$$

where the final term is nonpositive under MTSn. This implies that $P\{Y(0)=1\} \leq P(Y=1|D^*=0)$. The UB for $P\{Y(0)=1\}$ is therefore obtained assuming $P\{Y(1)=1|D^*=0\}=P(Y=1|D^*=1)$, and the LB is obtained assuming $P\{Y(0)=1|D^*=1\}=0$. This yields

$$P\{Y(0)=1\} \in \left\{P(Y=1,D^*=0), \frac{P(Y=1,D^*=0)}{P(D^*=0)}\right\}$$

Allowing for measurement error, these bounds become

$$\begin{aligned}P\{Y(1)=1\} &\in \left\{\frac{P(Y=1,D=1)-\theta_1^+ +\theta_1^-}{P(D=1)-(\theta_1^+ +\theta_0^+)+(\theta_1^- +\theta_0^-)},\right.\\ &\quad \left. P(D=0)+P(Y=1,D=1)+\theta_0^+ -\theta_0^-\right\}\\ &\equiv (\text{LB}_1, \text{UB}_1)\\ P\{Y(0)=1\} &\in \left\{P(Y=1,D=0)+\theta_1^+ -\theta_1^-, \frac{P(Y=1,D=0)+\theta_1^+ -\theta_1^-}{P(D=0)+(\theta_1^+ +\theta_0^+)-(\theta_1^- +\theta_0^-)}\right\}\\ &\equiv (\text{LB}_0, \text{UB}_0)\end{aligned}$$

Under assumption A1 above [arbitrary errors with UB, $P(Z^*=0) \leq Q$], the LB is given in (8) and the UB in (10). Similarly, under assumption A2 [no false positives, $P(Z^*|D=1)=1$], the LB is given in (9) and the UB in (11).

Positive monotone treatment selection (MTSp)

MTSp refers to the case of positive selection. In this case, individuals in the treatment group are more likely to experience a good outcome conditional on treatment assignment. Being that $Y = 1$ denotes a good outcome, MTSp implies that

$$P\{Y(1) = 1|D^* = 1\} \geq P\{Y(1) = 1|D^* = 0\}$$
$$P\{Y(0) = 1|D^* = 1\} \geq P\{Y(0) = 1|D^* = 0\}$$

Similar to the discussion on MTSn, we can rewrite $P\{Y(1) = 1\}$ as

$$P\{Y(1) = 1\} = P(Y = 1|D^* = 1) + P(D^* = 0)[P\{Y(1) = 1|D^* = 0\} - P(Y = 1|D^* = 1)]$$

where the final term is nonpositive under MTSp. This implies that $P\{Y(1) = 1\} \leq P(Y = 1|D^* = 1)$. The LB for $P\{Y(1) = 1\}$ is obtained assuming $P\{Y(1) = 1|D^* = 0\} = 0$, and the UB is obtained assuming $P\{Y(1) = 1|D^* = 0\} = P(Y = 1|D^* = 1)$. This yields

$$P\{Y(1) = 1\} \in \left\{P(Y = 1, D^* = 1), \frac{P(Y = 1, D^* = 1)}{P(D^* = 1)}\right\}$$

A similar inspection of $P\{Y(0) = 1\}$ yields

$$P\{Y(0) = 1\} = P(Y = 1|D^* = 0) + P(D^* = 1)[P\{Y(0) = 1|D^* = 1\} - P\{Y(0) = 1|D^* = 0\}]$$

where the final term is nonnegative under MTSp. This implies that $P\{Y(0) = 1\} \geq P(Y = 1|D^* = 0)$. The LB for $P\{Y(0) = 1\}$ is obtained assuming $P\{Y(1) = 1|D^* = 0\} = P(Y = 1|D^* = 1)$, and the UB is obtained assuming $P\{Y(0) = 1|D^* = 1\} = 1$. This yields

$$P\{Y(0) = 1\} \in \left\{\frac{P(Y = 1, D^* = 0)}{P(D^* = 0)}, P(Y = 1, D^* = 0) + P(D^* = 1)\right\}$$

Allowing for measurement error, these bounds become

$$P\{Y(1) = 1\} \in \left\{P(Y = 1, D = 1) + \theta_1^- - \theta_1^+, \frac{P(Y = 1, D = 1) + \theta_1^- - \theta_1^+}{P(D = 1) - (\theta_1^+ + \theta_0^+) + (\theta_1^- + \theta_0^-)}\right\} \equiv (\mathrm{LB}_1, \mathrm{UB}_1)$$

$$P\{Y(0) = 1\} \in \left\{\frac{P(Y = 1, D = 0) + \theta_1^+ - \theta_1^-}{P(D = 0) + (\theta_1^+ + \theta_0^+) - (\theta_1^- + \theta_0^-)} P(Y = 1, D = 0) + P(D = 1) + \theta_0^- - \theta_0^+\right\} \equiv (\mathrm{LB}_0, \mathrm{UB}_0)$$

Under assumption A1 above [arbitrary errors with UB, $P(Z^* = 0) \leq Q$], the LB is given in (10) and the UB in (8). Similarly, under assumption A2 [no false positives, $P(Z^*|D = 1) = 1$], the LB is given in (11) and the UB in (9).

2.5 MTR

The MTR assumption may be invoked in addition to MTS or in isolation. MTR assumes that individuals do not select into a treatment that would make them worse off in expectation. Again, one must be cautious in interpreting each case depending on whether the outcome, Y, is desirable. Here we assume that $Y = 1$ corresponds to the desirable outcome. Thus the MTR assumption implies that $P\{Y(1) = 1\} \geq P\{Y(0) = 1\}$.

Invoking MTR in isolation implies that the UBs are given by the worst-case UBs in (10) and (11) under the assumptions of arbitrary errors and no false positives, respectively. The worst-case LBs, however, are now replaced by 0. Combining MTR and MTSn (or MTSp) implies that the UBs are given by the UBs under MTSn (MTSp), while the LBs again are replaced by 0.

2.6 MIV

The MIV assumption implies that the latent probability of a good outcome conditional on treatment assignment—$P\{Y(1) = 1\}$ and $P\{Y(0) = 1\}$—varies (weakly) monotonically with an observed covariate. The MIV assumption alone has no identifying power; therefore, following Kreider and Pepper (2007) and Kreider et al. (2012), we combine the MIV assumption with the MTS assumption and then, in the next section, with the MTS and MTR assumptions.

MIV and MTS

Let ν denote the monotone instrument and assume without loss of generality that $P\{Y(1) = 1\}$ and $P\{Y(0) = 1\}$ are nondecreasing in ν. Defining $u_1 < u < u_2$, the MIV assumption implies that

$$
\begin{aligned}
P\{Y(1) = 1|\nu = u_1\} \leq \quad & P\{Y(1) = 1|\nu = u\} \quad \leq P\{Y(1) = 1|\nu = u_2\} \\
P\{Y(0) = 1|\nu = u_1\} \leq \quad & P\{Y(0) = 1|\nu = u\} \quad \leq P\{Y(0) = 1|\nu = u_2\}
\end{aligned}
$$

Recall, the MTSn assumption implies that

$$
\begin{aligned}
P\{Y(1) = 1|D^* = 1\} \leq P\{Y(1) = 1|D^* = 0\} \\
P\{Y(0) = 1|D^* = 1\} \leq P\{Y(0) = 1|D^* = 0\}
\end{aligned}
$$

such that without measurement error the bounds are given by

$$
\begin{aligned}
P\left\{Y(1) = 1\right\} &\in \left\{\frac{P\left(Y = 1, D^* = 1\right)}{P(D^* = 1)}, P(D^* = 0) + P\left(Y = 1, D^* = 1\right)\right\} \\
P\left\{Y(0) = 1\right\} &\in \left\{P\left(Y = 1, D^* = 0\right), \frac{P\left(Y = 1, D^* = 0\right)}{P(D^* = 0)}\right\}
\end{aligned}
$$

Combining MIV and MTSn involves the following steps:[3]

1. Split the sample into J cells, $j = 1, 2, \ldots, J$, based on values of ν, and let P_j denote the sample fraction in cell j.

2. Calculate the MTSn bounds for $P\{Y(1) = 1\}$ and $P\{Y(0) = 1\}$ for each cell under assumption A1 or A1 and A2. This yields UB_d^j and LB_d^j, $j = 1, 2, \ldots, J$ and $d = 0, 1$.

3. Calculate the overall LB for $P\{Y(1) = 1\}$, denoted by LB_1, as[4]

$$\text{LB}_1 = T_n = \sum_j P_j \left(\sup_{j' \leq j} \text{LB}_1^{j'} \right)$$

   T_n is therefore a weighted average of the appropriate LB estimates across all J cells, constructed by taking LB_1 in each jth cell and averaging over the supremum of the individual LBs across all cells below and including the jth cell.

4. Adjust LB_1 for finite-sample bias by bootstrapping the sampling distribution of LB_1^j.

   a. Randomly draw with replacement K independent pseudosamples from the original data of size N with replacement.

   b. Compute LB_{1k}^j, $j = 1, 2, \ldots, J$ and $k = 1, 2, \ldots, K$.

   c. Compute $T_n^k = \sum_j P_{jk}(\sup_{j' \leq j} \text{LB}_{1k}^{j'})$, $k = 1, 2, \ldots, K$. This process is identical to that in step 3 above, performed separately for each kth pseudo-sample.

   d. Compute the sample mean from the bootstrap as

$$E^*(T_n) = \frac{1}{K} \sum_k T_n^k$$

   e. The estimated bias is given by

$$\widehat{b} = E^*(T_n) - T_n$$

   f. The bias-corrected LB is given by

$$\text{LB}_1 = T_n^c = T_n - \widehat{b} = 2T_n - E^*(T_n)$$

5. Calculate the overall UB for $P\{Y(1) = 1\}$, denoted by UB_1, as

$$\text{UB}_1 = U_n = \sum_j P_j \left(\inf_{j' \geq j} \text{UB}_1^{j'} \right)$$

---

3. Combining MIV with MTSp follows similarly and is therefore excluded for brevity.
4. As before, one must be cautious because this assumes $Y = 1$ is a desirable outcome.

This assumes $Y(1) = 1$ and $Y(0) = 1$ are good outcomes. If $Y(1) = 1$ and $Y(0) = 1$ are bad outcomes, then $\text{UB}_1 = U_n = \sum_j P_j(\inf_{j' \leq j} \text{UB}_1^{j'})$. As with the LB in step 3, T_n is again a weighted average of the appropriate UB estimates across all J cells, constructed by taking UB_1 in each jth cell and averaging over the infimum of the individual UBs across all cells above and including the jth cell.

6. Adjust UB_1 for finite-sample bias by bootstrapping the sampling distribution of UB_1^j.

   a. Randomly draw with replacement K independent pseudosamples from the original data of size N with replacement.

   b. Compute UB_{1k}^j, $j = 1, 2, \ldots, J$ and $k = 1, 2, \ldots, K$.

   c. Compute $U_n^k = \sum_j P_{jk}(\inf_{j' \geq j} \text{UB}_{1k}^{j'})$, $k = 1, 2, \ldots, K$. This process is identical to that in step 5 above, performed separately for each kth pseudosample.

   d. Compute the sample mean from the bootstrap as

$$E^*(U_n) = \frac{1}{K} \sum_k U_n^k$$

   e. The estimated bias is given by

$$\widehat{b} = E^*(U_n) - U_n$$

   f. The bias-corrected UB is given by

$$\text{UB}_1 = U_n^c = U_n - \widehat{b} = 2U_n - E^*(U_n)$$

7. Repeat steps 3–6 to obtain the overall LB and UB for $P\{Y(0) = 1\}$, denoted LB_0 and UB_0, respectively.

8. Obtain bounds for the ATE given by

$$\text{ATE} \in (\text{LB}_1 - \text{UB}_0, \text{UB}_1 - \text{LB}_0)$$

Prior to continuing, two comments are necessary. Firstly, the MIV estimator suffers from finite-sample bias (Manski and Pepper 2000). Steps 4 and 6 in the preceding-algorithm follow Kreider et al. (2012) and use the nonparametric finite-sample bias-corrected MIV estimator put forth in Kreider and Pepper (2007). However, Hirano and Porter (2012) caution against the use of bias-corrected techniques because such procedures cannot fully eliminate the bias in the case of nonsmooth estimators and may cause substantial increases in variance. Thus, users of `tebounds` may wish to assess the sensitivity of the bounds to the use of the bias correction.

An alternative to the bias correction procedure used in Kreider et al. (2012) is the precision-corrected approach recently proposed in Chernozhukov, Lee, and Rosen (2013).[5] This procedure adjusts the terms $\text{LB}^{j'}$ and $\text{UB}^{j'}$ in steps 3 and 6, respectively, before taking the sup or inf. Thus, the correction is applied during the estimation of the bounds of the conditional probabilities, $P\{Y(d) = 1|\nu = u\}$, $d = 0, 1$. In contrast, the approach in Kreider et al. (2012) computes the bounds for each MIV cell, indexed by j, then constructs the weighted averages of the LBs and UBs across the different MIV cells, and finally applies the finite-sample correction to the estimated bounds of the unconditional probabilities, $P\{Y(d) = 1\}$, $d = 0, 1$.

Here we follow the Kreider et al. (2012) approach for two reasons. First, it is computationally simpler. Second, and more importantly, Chernozhukov, Lee, and Rosen (2013) discuss only the estimated bounds of the conditional probabilities, $P\{Y(d) = 1|\nu = u\}$, $d = 0, 1$, and the associated inference. It is not obvious how this approach should be extended when the focus is on estimation and inferences of the bounds on the ATE.

Secondly, the asymptotic properties of estimators involving nonsmooth functions like sup and inf, such as those based on an MIV, are the subject of recent debate. Manski and Pepper (2000, 1007) note that "the consistency of the resulting bounds estimates is easy to establish". However, Hirano and Porter (2012, 1769) show that in such cases "no locally asymptotically unbiased estimators exist". The distinction may lie in that the objects of interest in Manski and Pepper (2000), and here, are the bounds on the ATE. As a result, the weighted averages utilized in steps 3 and 5 yield a smooth estimator. In any event, obviously the asymptotic properties of such estimators have important implications for conducting proper inference. Because this is the subject of ongoing research (see, for example, Chernozhukov, Lee, and Rosen [2013] and the references therein), users of the `tebounds` should keep abreast of developments in the literature.

2.7 MIV, MTR, and MTS

Combining the MIV, MTR, and MTS assumptions can further tighten the ATE bounds. The addition of the MTR assumption within each cell of the MIV is, however, a bit different than the imposition of the MTR (or MTS + MTR) assumption discussed previously. The difference arises because implementation of the MIV assumption does not entail bounding the ATE within each MIV cell. Rather, we are bounding the components of the ATE within each cell (that is, $P\{Y(1) = 1\}$ and $P\{Y(0) = 1\}$ separately, rather than $P\{Y(1) = 1\} - P\{Y(0) = 1\}$). Because the MTR assumption requires $P\{Y(1) = 1\} - P\{Y(0) = 1\} \geq 0$, this implies that, within each MIV cell, the LB of $P\{Y(1) = 1\}$ cannot be strictly less than the UB of $P\{Y(0) = 1\}$.

5. See the author-provided command `clrbounds`.

3 The tebounds command

3.1 Syntax

The `tebounds` command estimates bounds on the ATE under the assumptions discussed above. The syntax for the `tebounds` command is

`tebounds` *depvar* [*if*] [*in*] [*weight*], `treat(`*varname*`)` [`control(`#`)` `treatment(`#`)` `miv(`*varname*`)` `ncells(`#`)` `erates(`*string*`)` `k(`#`)` `np(`#`)` `bs` `reps(`#`)` `level(`#`)` `im` `survey` `weights(`#`)` `npsu(`#`)` `nodisplay` `graph` `saving(`*filename*`)` `replace`]

The `tebounds` command requires that `bsweights` (Kolenikov 2010) and `bs4rw` (Pitblado 2013) also be installed.

3.2 Options

`treat(`*varname*`)` specifies the variable name of the treatment indicator. `treat()` is required.

`control(`#`)` specifies the numeric value of `treat()` used to indicate the control group. The default is `control(0)`.

`treatment(`#`)` specifies the numeric value of `treat()` used to indicate the treatment group. The default is `treatment(1)`.

`miv(`*varname*`)` specifies the MIV name. The use of `miv(`*varname*`)` also calls a secondary MIV command from within the `tebounds` command, the results of which are passed into the larger `tebounds` command. See Manski and Pepper (2000), Kreider and Pepper (2007), Kreider et al. (2012), Manski and Pepper (2009), and Lafférs (2013) for a more detailed definition and discussion of proper MIVs.

`ncells(`#`)` denotes the number of cells used in the MIV estimator. The default is `ncells(5)`. The MIV variable is divided by percentiles according to the number of cells specified in `ncells()`. For example, `ncells(5)` will split the MIV variable into five groups according to the 20th, 40th, 60th, 80th, and 100th percentile values.

`erates(`*string*`)` denotes the assumed rates of measurement error used to identify the ATE bounds. The default is `erates(0 0.05 0.10 0.25)` indicating assumed rates of 0%, 5%, 10%, and 25% measurement error.

`k(`#`)` denotes the number of bootstrap replications used for the MIV bias correction. The default is `k(100)`.

`np(`#`)` denotes the number of intervals over which the grid search is performed to tighten the bounds in the arbitrary classification-error model (Kreider, Pepper, and Roy 2013).

`bs`, `reps(#)`, and `level(#)` specify that confidence intervals (CIs) be calculated, as a percentage specified in `level()` (default is `level(95)`), by bootstrap (`bs`) based on the number of replications in `reps(#)`. The default is `reps(100)`.

`im` specifies that CIs be calculated following Imbens and Manski (2004). If left unspecified, CIs are calculated using the percentile method.

`survey` indicates that data are survey data and that survey weights have been assigned with `svyset`. The survey weights will be used as part of the ATE calculation as well as the bootstrap replications, if relevant. If bootstrap CIs are also requested, the `tebounds` command first estimates bootstrap weights via the `bsweights` command. Bootstrap weights will also be calculated as part of the MIV bias correction. Attempts to call the `survey` option will result in an error if the user has not first declared the survey design with `svyset`.

`weights(#)` indicates whether bootstrap weights have already been calculated. The default is `weights(0)`, which indicates that bootstrap weights have not been calculated and will instead be calculated from within the `tebounds` command. The `weights()` option is intended as a programming control to avoid replacing the estimates in `ereturn` at each iteration of the bootstrap. Should the user have prespecified bootstrap weights available, the weights must be of the form `bsw1 bsw2` ... `bsw`*N*, where *N* denotes the number of bootstrap replications.[6]

`npsu(#)` specifies the value of `n(#)` in the `bsweights` command. `npsu()` specifies how the number of primary sampling units per stratum are handled. The default is `npsu(-1)`, which indicates a bootstrap sample of size $n_h - 1$ for all strata h (Kolenikov 2010).

`nodisplay` suppresses the summary results table. Results are still stored in `ereturn`.

`graph` specifies that ATE bounds be graphed as a function of the maximum rates of measurement error from `erates`.

`saving(`*filename*`)` indicates the location in which to save the output.

`replace` indicates that the output in `saving()` should replace any preexisting file in the same location.

4 Examples

4.1 U.S. School Breakfast Program

Following Millimet and Tchernis (2013), we provide an application of the `tebounds` command to the study of the U.S. School Breakfast Program (SBP). Specifically, we seek bounds for the ATEs of SBP on child weight. The data are from the Early Childhood

6. The `weights()` option considers only the non-MIV estimators. The bootstrap weights in the MIV estimators must be calculated from within the command itself because these weights are dependent on the MIV variable as well as the number of cells in `ncells()`.

Longitudinal Study, Kindergarten Class of 1998–99 and are available for download from the *Journal of Applied Econometrics* data archive.[7] We partially identify the ATE of self-reported SBP participation in first grade (`break1`) on the probability of being not obese in spring of third grade (`NOTobese`).

In our application, we allow the maximum misclassification rate of program participation, Q, to be 0%, 1%, 2%, 5%, and 10%. We use an index of socioeconomic status (`ses`) as the MIV and divide the sample into 20 cells. Bootstrap CIs at 95% based on 100 replications are provided. The resulting output is as follows:

```
. use mmrexampledata
. tebounds NOTobese, treat(break1) erates(0 1 2 5 10) ncells(20) miv(ses) bs
> reps(100) graph saving(SJgraphsSBP) replace
----------------------------------------------------------------------------
Outcome:  NOTobese
Treatment: break1
Number of pseudo-samples used in MIV bias correction: 100
Number of bootstrap reps for 95% CIs: 100
----------------------------------------------------------------------------
 Error Rate     | Arbitrary Errors                 | No False Positives
----------------------------------------------------------------------------
Exogenous Selection Model
  0             [ -0.059, -0.059] p.e.           [ -0.059, -0.059] p.e.
                [ -0.076, -0.040] CI             [ -0.076, -0.040] CI
  .01           [ -0.095, -0.022] p.e.           [ -0.094, -0.049] p.e.
                [ -0.112, -0.002] CI             [ -0.111, -0.030] CI
  .02           [ -0.130,  0.017] p.e.           [ -0.129, -0.039] p.e.
                [ -0.146,  0.037] CI             [ -0.146, -0.021] CI
  .05           [ -0.229,  0.144] p.e.           [ -0.229, -0.012] p.e.
                [ -0.246,  0.165] CI             [ -0.245,  0.005] CI
  .1            [ -0.384,  0.267] p.e.           [ -0.384,  0.030] p.e.
                [ -0.399,  0.276] CI             [ -0.399,  0.045] CI
----------------------------------------------------------------------------
No Monotonicity Assumptions (Worst Case Selection)
  0             [ -0.641,  0.359] p.e.           [ -0.641,  0.359] p.e.
                [ -0.650,  0.369] CI             [ -0.650,  0.369] CI
  .01           [ -0.651,  0.369] p.e.           [ -0.651,  0.369] p.e.
                [ -0.660,  0.379] CI             [ -0.660,  0.379] CI
  .02           [ -0.661,  0.379] p.e.           [ -0.661,  0.379] p.e.
                [ -0.670,  0.389] CI             [ -0.670,  0.389] CI
  .05           [ -0.691,  0.409] p.e.           [ -0.691,  0.409] p.e.
                [ -0.700,  0.419] CI             [ -0.700,  0.419] CI
  .1            [ -0.741,  0.459] p.e.           [ -0.741,  0.459] p.e.
                [ -0.750,  0.469] CI             [ -0.750,  0.469] CI
----------------------------------------------------------------------------
```

7. See http://qed.econ.queensu.ca/jae/datasets/millimet001/.

Error Rate	Arbitrary Errors		No False Positives	
MTS Assumption: Negative Selection				
0	[-0.059, 0.359]	p.e.	[-0.059, 0.359]	p.e.
	[-0.076, 0.369]	CI	[-0.076, 0.369]	CI
.01	[-0.095, 0.369]	p.e.	[-0.094, 0.369]	p.e.
	[-0.112, 0.379]	CI	[-0.111, 0.379]	CI
.02	[-0.130, 0.379]	p.e.	[-0.129, 0.379]	p.e.
	[-0.146, 0.389]	CI	[-0.146, 0.389]	CI
.05	[-0.229, 0.409]	p.e.	[-0.229, 0.409]	p.e.
	[-0.246, 0.419]	CI	[-0.245, 0.419]	CI
.1	[-0.384, 0.459]	p.e.	[-0.384, 0.459]	p.e.
	[-0.399, 0.469]	CI	[-0.399, 0.469]	CI
MTS and MTR Assumptions: Negative Selection				
0	[0.000, 0.359]	p.e.	[0.000, 0.359]	p.e.
	[0.000, 0.369]	CI	[0.000, 0.369]	CI
.01	[0.000, 0.369]	p.e.	[0.000, 0.369]	p.e.
	[0.000, 0.379]	CI	[0.000, 0.379]	CI
.02	[0.000, 0.379]	p.e.	[0.000, 0.379]	p.e.
	[0.000, 0.389]	CI	[0.000, 0.389]	CI
.05	[0.000, 0.409]	p.e.	[0.000, 0.409]	p.e.
	[0.000, 0.419]	CI	[0.000, 0.419]	CI
.1	[0.000, 0.459]	p.e.	[0.000, 0.459]	p.e.
	[0.000, 0.469]	CI	[0.000, 0.469]	CI
MIV and MTS Assumptions: Negative Selection				
0	[0.044, 0.352]	p.e.	[0.044, 0.352]	p.e.
	[0.008, 0.358]	CI	[0.008, 0.358]	CI
.01	[0.003, 0.369]	p.e.	[0.003, 0.362]	p.e.
	[-0.032, 0.378]	CI	[-0.032, 0.368]	CI
.02	[-0.036, 0.379]	p.e.	[-0.036, 0.372]	p.e.
	[-0.072, 0.389]	CI	[-0.072, 0.378]	CI
.05	[-0.123, 0.409]	p.e.	[-0.123, 0.402]	p.e.
	[-0.160, 0.419]	CI	[-0.160, 0.408]	CI
.1	[-0.224, 0.459]	p.e.	[-0.212, 0.452]	p.e.
	[-0.248, 0.469]	CI	[-0.249, 0.458]	CI
MIV, MTS, MTR Assumptions: Negative Selection				
0	[0.044, 0.352]	p.e.	[0.044, 0.352]	p.e.
	[0.008, 0.358]	CI	[0.008, 0.358]	CI
.01	[0.003, 0.369]	p.e.	[0.003, 0.362]	p.e.
	[0.000, 0.378]	CI	[0.000, 0.368]	CI
.02	[0.000, 0.379]	p.e.	[0.000, 0.372]	p.e.
	[0.000, 0.389]	CI	[0.000, 0.378]	CI
.05	[0.000, 0.409]	p.e.	[0.000, 0.402]	p.e.
	[0.000, 0.419]	CI	[0.000, 0.408]	CI
.1	[0.000, 0.459]	p.e.	[0.000, 0.452]	p.e.
	[0.000, 0.469]	CI	[0.000, 0.458]	CI

As indicated by the section headings, the output presents the point estimates (`p.e.`) for the bounds for the ATE, as well as 95% CIs (`CI`) under each set of assumptions concerning the nature of the selection process and the type of misclassification (arbitrary errors or no false positives). Moreover, within each panel, separate bounds are presented for each value of Q specified in the `tebounds` command (in this case, 0%, 1%, 2%, 5%, and 10%). A selection of the graphs produced are illustrated in figures 1–3.

In terms of the bounds, we see that the mean difference in outcomes across the treatment and control groups, assuming SBP participation is not misreported, is -0.059. This implies that participants in the SBP are 5.9% more likely to be obese relative to nonparticipants. However, because the SBP is subsidized for low-income households, and low-income individuals are more likely to be obese, the assumption of exogenous selection is not reasonable. That said, we also see that the assumption of exogenous selection is insufficient to identify the sign of the ATE if at least 2% (10%) of the sample misreport program participation under the assumption of arbitrary errors (no false positives). This result illustrates two important facts. First, the association between SBP participation and child weight is not robust to even small amounts of misreporting. Second, the assumption of no false positives provides some identifying information relative to the assumption of arbitrary errors.

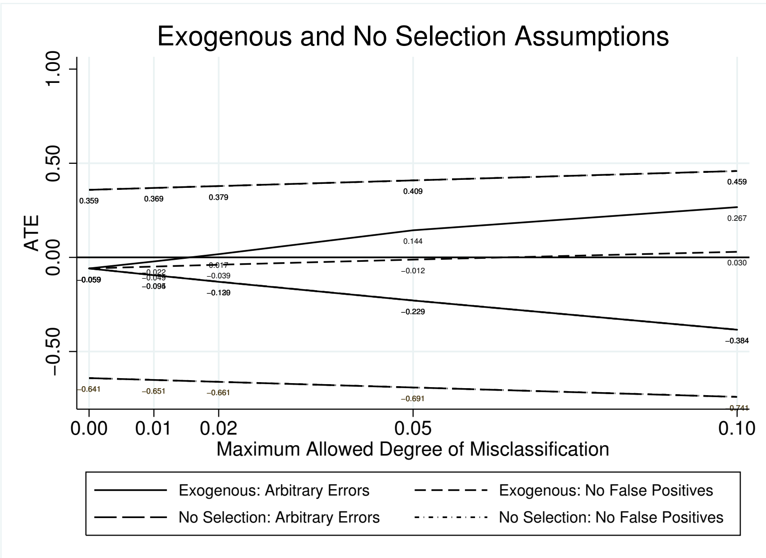

Figure 1. Bounds under exogenous selection and worst case

The worst-case bounds, assuming no misclassification, have a width of unity and necessarily include 0. Admitting the possibility of misreporting simply widens the bounds. However, these bounds still include some large values for the ATE that may be useful to policymakers. For example, if the SBP program would pass (fail) a cost–benefit analysis despite assuming the ATE on child weight is -0.641 (0.359), then precise knowledge of the ATE is unnecessary.

For the bounds invoking the monotonicity assumptions, we only display the bounds assuming negative selection into the SBP in the interest of brevity. Under MTSn alone, we see the LB comes from the bounds obtained under exogenous selection, while the UB comes from the worst-case bounds. Thus the width of the bounds shrink relative to the worst-case bounds but still fail to identify the sign of the ATE. Under MTSn and MTR, the LB is replaced by 0.[8] Thus the bounds still fail to identify the sign of the ATE.

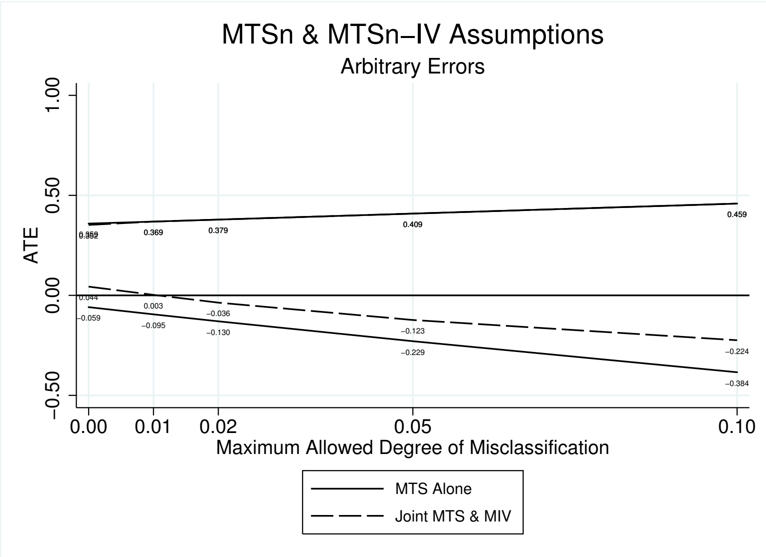

Figure 2. Bounds under MTSn and MIV + MTSn with arbitrary errors

The final two panels (figures 2 and 3) use the MIV assumption in addition to MTSn or MTSn and MTR. In both cases, we see that the bounds are strictly positive assuming that no more than 1% of the sample misreports treatment assignment. In these cases, we can conclude under fairly innocuous assumptions that participation in the SBP has a positive, causal effect on the probability of an average child being nonobese. This is an important policy finding and is consistent with alternative estimators discussed in Millimet and Tchernis (2013) that point identify the ATE under stringent assumptions. That said, the sensitivity of the bounds to classification errors is noteworthy and highlights the econometric importance of even relatively infrequent misreporting.

8. Arguably, the MTR assumption may be difficult to justify in the current application. We display this case purely for illustrative purposes.

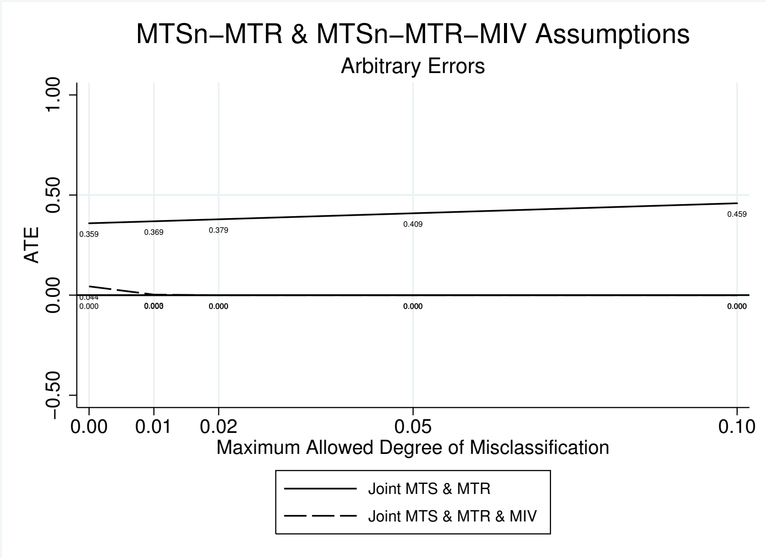

Figure 3. Bounds under MTSn + MTR and MIV + MTSn + MTR with arbitrary errors

4.2 Monte Carlo

To further illustrate the `tebounds` command, we undertake a simple Monte Carlo exercise. We simulate 10,000 datasets with 2,000 observations from the following data-generating process:

$$
\begin{aligned}
y_0 &= I\{\nu_0 + 2(z-\epsilon)\} \\
y_1 &= I\{\nu_1 + 2(z-\epsilon)\} \\
D^* &= I(2*\epsilon + \nu_2) \\
D &= \begin{cases} 1 - D^* & \text{if } U < 0.1 \\ D^* & \text{otherwise} \end{cases} \\
\nu_0 &\sim N(-1, 0.1) \\
\nu_1 &\sim N(-2, 0.1) \\
\nu_2 &\sim N(0, 1) \\
\epsilon &\sim N(0, 1) \\
z &\sim N(0, 1) \\
U &\sim U(0, 1)
\end{aligned}
$$

The population ATE is approximately 0.4. z is a valid MIV because both potential outcomes are increasing in z. The presence of ϵ leads to negative selection into the treatment. The misclassification rate is 10% and is arbitrary.

The results are summarized in tables 1 and 2, where the MIV bounds are obtained using 20 cells. Table 1 reports the coverage rates (that is, the fraction of simulations where the bounds encompass the true value of the ATE). Not surprisingly, the exogenous selection bounds fail to cover the true value even when a maximum misclassification rate of 10% (the true value in the population) is allowed. In all other cases, the bounds always include the true value of the ATE, even when the incorrect assumption of no misclassification is imposed.

Table 1. Coverage rates

Assumption	Q	Coverage rate
exogenous selection	0	0.000
	10	0.000
worst case	0	1.000
	10	1.000
MTSn	0	1.000
	10	1.000
MTSn + MTR	0	1.000
	10	1.000
MIV + MTSn	0	1.000
	10	1.000
MIV + MTSn + MTR	0	1.000
	10	1.000

Table 2 focuses on one aspect of the information that is potentially learned from the partial identification approach: the sign of the ATE. Specifically, table 2 reports the fraction of simulations where the bounds are strictly positive (that is, able to exclude 0 and produce the correct sign of the ATE). Focusing on the bounds that utilize the various monotonicity assumptions, we see that the bounds obtained with (without) the MIV assumption are able to correctly sign the ATE in 95% (89%) of the simulations under the assumption of no misclassification. However, when we allow for the true rate of misclassification in the population, the bounds always include 0. It is important to note that while the bounds obtained under monotonicity and the assumption of no misclassification exclude 0 in the majority of simulations and always contain the true value of the ATE (see table 1), there is nothing that guarantees this will always be the case. Thus, one should avoid drawing the conclusion from this simple illustration that it is acceptable to focus on the bounds obtained under the assumption of no misclassification when, in fact, misreporting is a feature of the data.

Table 2. Frequency of bounds excluding 0

Assumption	Q	Percentage
exogenous selection	0	0.890
	10	0.000
worst case	0	0.000
	10	0.000
MTSn	0	0.890
	10	0.000
MTSn + MTR	0	0.890
	10	0.000
MIV + MTSn	0	0.950
	10	0.000
MIV + MTSn + MTR	0	0.950
	10	0.000

5 Remarks

The `tebounds` command provides a means to partially identify the ATE of a binary treatment on a binary outcome under a variety of assumptions concerning the nature of the self-selection process and the nature and frequency of misreporting of treatment assignment. The binary outcome should be defined such that $Y = 1$ corresponds to the desirable outcome. CIs are available following Imbens and Manski (2004).

6 References

Bontemps, C., T. Magnac, and E. Maurin. 2012. Set identified linear models. *Econometrica* 80: 1129–1155.

Chernozhukov, V., S. Lee, and A. M. Rosen. 2013. Intersection bounds: Estimation and inference. *Econometrica* 81: 667–737.

Gundersen, C., B. Kreider, and J. Pepper. 2012. The impact of the National School Lunch Program on child health: A nonparametric bounds analysis. *Journal of Econometrics* 166: 79–91.

Hirano, K., and J. R. Porter. 2012. Impossibility results for nondifferentiable functionals. *Econometrica* 80: 1769–1790.

Imbens, G. W., and C. F. Manski. 2004. Confidence intervals for partially identified parameters. *Econometrica* 72: 1845–1857.

Kolenikov, S. 2010. Resampling variance estimation for complex survey data. *Stata Journal* 10: 165–199.

Kreider, B., and J. V. Pepper. 2007. Disability and employment: Reevaluating the evidence in light of reporting errors. *Journal of the American Statistical Association* 102: 432–441.

Kreider, B., J. V. Pepper, C. Gundersen, and D. Jolliffe. 2012. Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported. *Journal of the American Statistical Association* 107: 958–975.

Kreider, B., J. V. Pepper, and M. Roy. 2013. Does the women, infants, and children (WIC) program improve infant health outcomes? Unpublished manuscript.

Lafférs, L. 2013. A note on bounding average treatment effects. *Economics Letters* 120: 424–428.

Manski, C. F. 2013. *Public Policy in an Uncertain World: Analysis and Decisions*. Cambridge, MA: Harvard University Press.

Manski, C. F., and J. V. Pepper. 2000. Monotone instrumental variables: With an application to the returns to schooling. *Econometrica* 68: 997–1010.

———. 2009. More on monotone instrumental variables. *Econometrics Journal* 12: S200–S216.

Millimet, D. L. 2011. The elephant in the corner: A cautionary tale about measurement error in treatment effects models. *Advances in Econometrics: Missing-Data Methods* 27A: 1–39.

Millimet, D. L., and R. Tchernis. 2013. Estimation of treatment effects without an exclusion restriction: With an application to the analysis of the school breakfast program. *Journal of Applied Econometrics* 28: 982–1017.

Pitblado, J. 2013. bs4rw: Bootstrap using replicate weight variables. http://www.stata.com/users/jpitblado.

Tamer, E. 2010. Partial identification in econometrics. *Annual Review of Economics* 2: 167–195.

About the authors

Ian McCarthy is an assistant professor of economics at Emory University. His research relates primarily to the fields of health economics, policy, and economic evaluation of health care programs. Within these areas, he is interested in patient choice, hospital and insurance market structure, and empirical methodologies in cost and comparative effectiveness research. Prior to joining Emory University, he was a director in the economic consulting practice at FTI Consulting and a director of health economics with Baylor Scott & White Health. He received his PhD in economics from Indiana University.

Daniel Millimet is a professor of economics at Southern Methodist University and a research fellow at the Institute for the Study of Labor. His primary areas of research are applied microeconometrics, labor economics, and environmental economics. His research has been funded by various organizations, including the United States Department of Agriculture. He received his PhD in economics from Brown University.

Manan Roy is a postdoctoral scholar at the Department of Public Policy, University of North Carolina at Chapel Hill. Her primary areas of research are applied microeconometrics, program evaluation, health economics, and labor economics. Within these areas, she is interested in food insecurity, food and nutrition programs, childhood obesity, and measurement error. She received her PhD in economics from Southern Methodist University.