Bounding Treatment Effects: Stata 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
IZA, Bonn, Germany
millimet@smu.edu

Manan Roy
IMPAQ International, Columbia, Maryland
mroy@impaqint.com

Abstract. We present a new Stata command, tebounds, which implements a variety of techniques to bound the average treatment effect (ATE) 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 (MTS), Monotone Treatment Response (MTR), and Monotone Instrumental Variable (MIV) assumptions of Manski and Pepper (2000), Kreider and Pepper (2007), Kreider et al. (2012) and Gundersen et al. (2012) may be imposed. Imbens-Manski confidence intervals are provided.

Keywords: st0001, tebounds.ado, Treatment Effects, Selection, Misreporting, Monotone Instrumental Variable, Monotone Treatment Selection, Monotone Treatment Response

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), there exist a variety of appropriate methodologies 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 in order to achieve point identifi-
cation. 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 as they highlight what may be learned without invoking perhaps untenable assumptions. As Manski (2013, p. 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 et al. (2012, p. 1129) write: “Point identification is often achieved by using strong and difficult to motivate restrictions on the parameters of interest.” Tamer (2010, p. 168) states: “Stronger assumptions will lead to more information about a parameter, but less credible inferences can be conducted.” Instead, Manski (2013, p. 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, p. 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 zero, thereby identifying the sign of the parameter. Moreover, bounds can exclude extreme values which may, for example, be useful in determining if 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

\[
ATE(1, 0) = P[Y(1) = 1 | X \in \Omega] - P[Y(0) = 1 | X \in \Omega]
\]  

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 since 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 whose values lie in the set \(\Omega\). To simplify notation, the conditioning on \(X\) is left implicit throughout the remainder of the paper.

---

\(^1\)Bontemps et al. (2012, p. 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.”
The terms in (1) can be written as:

\[
P[Y(1) = 1] = P[Y(1) = 1|D^* = 1] P(D^* = 1) \\
+ P[Y(1) = 1|D^* = 0] P(D^* = 0)
\]  

(2)

\[
P[Y(0) = 1] = P[Y(0) = 1|D^* = 1] P(D^* = 1) \\
+ P[Y(0) = 1|D^* = 0] P(D^* = 0)
\]

(3)

There are two problems confronted 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 \(P[Y(0) = 1|D^* = 0]\). However, the sampling process cannot identify the counterfactual probabilities, \(P[Y(1) = 1|D^* = 0]\) and \(P[Y(0) = 1|D^* = 1]\), so that \(P[Y(1) = 1]\) 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 or not, where \(Z^* = 1\) if \(D^* = D\) and zero otherwise. Third, define the following notation:

\[
\theta_1^+ \equiv P(Y = 1, D = 1, Z^* = 0) \Rightarrow \text{fraction of observations which are false positives with } Y = 1 \\
\theta_0^+ \equiv P(Y = 0, D = 1, Z^* = 0) \Rightarrow \text{fraction of observations which are false positives with } Y = 0 \\
\theta_1^- \equiv P(Y = 1, D = 0, Z^* = 0) \Rightarrow \text{fraction of observations which are false negatives with } Y = 1 \\
\theta_0^- \equiv P(Y = 0, D = 0, Z^* = 0) \Rightarrow \text{fraction of observations which are false negatives with } Y = 0.
\]
We can now further decompose \( P[Y(1) = 1] \) and \( P[Y(0) = 1] \) as follows:

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

\[
= [P(Y = 1, D = 1) - \theta_1^+ + \theta_1^-]
\]

\[
+ P[Y(1) = 1|D^* = 0] [P(D = 0) + (\theta_1^+ + \theta_0^-) - (\theta_1^- + \theta_0^-)] \tag{4}
\]

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

\[
= P[Y(0) = 1|D^* = 1] [P(D = 1) - (\theta_1^+ + \theta_0^-) + (\theta_1^- + \theta_0^-)]
\]

\[
+ P(Y = 1, D = 0) + \theta_1^+ - \theta_1^- \tag{5}
\]

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 in order 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 et al. (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:

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

\[
0 \leq \theta_1^- \leq \min\{Q, P(Y = 1, D = 0)\} = \theta_1^{UB^-}
\]

\[
0 \leq \theta_0^- \leq \min\{Q, P(Y = 0, D = 0)\} = \theta_0^{UB^-}
\]

\[
0 \leq \theta_1^+ \leq \min\{Q, P(Y = 1, D = 1)\} = \theta_1^{UB^+}
\]

\[
0 \leq \theta_0^+ \leq \min\{Q, P(Y = 0, D = 1)\} = \theta_0^{UB^+}
\]

\[
\theta_1^+ + \theta_1^- + \theta_0^+ + \theta_0^- \leq Q
\]
A2 No false positives, \( P(Z^* = 1|D = 1) = 1 \). This assumption simplifies equations (4) and (5) above since \( \theta_1^+ = \theta_0^+ = 0 \).

Assumptions regarding the selection process include:

1. Exogenous Selection;
2. Worst Case Selection (no assumption about selection);
3. Monotone Treatment Selection (MTS);
4. MTS and Monotone Treatment Response (MTR);
5. Monotone Instrumental Variable (MIV) and MTS; or
6. MIV+MTS+MTR.

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

2.2 Exogenous Selection Bounds

The assumption of exogenous selection implies

\[
P[Y(1) = 1] = P[Y(1) = 1|D^* = 1] = P[Y(1) = 1|D^* = 0] \quad \text{and}
\]

\[
P[Y(0) = 1] = P[Y(0) = 1|D^* = 1] = P[Y(0) = 1|D^* = 0].
\]

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

\[
P[Y(1) = 1] = P[Y = 1|D^* = 1] P(D^* = 1) + P[Y(1) = 1|D^* = 0] P(D^* = 0)
= P[Y = 1|D^* = 1] P(D^* = 1) + \left[ P[Y(1) = 1|D^* = 1] - P(D^* = 1) \right] + 1
= P[Y = 1|D^* = 1], \quad \text{and}
\]

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

respectively.

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

$$P[Y = 1|D^* = 1] = \frac{P[Y = 1, D^* = 1]}{P(D^* = 1)}$$

$$= \frac{P(Y = 1, D = 1) - \theta_1^+ + \theta_1^-}{P(D = 1) - (\theta_1^+ + \theta_0^+)} + (\theta_1^- + \theta_0^-)$$

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

$$= \frac{P(Y = 1, D = 0) + \theta_1^+ - \theta_1^-}{P(D = 0) + (\theta_1^+ + \theta_0^+)} - (\theta_1^- + \theta_0^-).$$

Thus, point identification is not possible except under strict assumptions about the values of the unknown quantities in equations (8) and (9). Instead, we may use Assumptions A1 and A2 to bound the ATE. For example, following Proposition 1 in Kreider and Pepper (2007), we know

$$P(Y = 1|D^* = 1) \in \left[ \frac{P(Y = 1, D = 1) - \delta}{P(D = 1) - 2\delta + Q}, \frac{P(Y = 1, D = 1) + \gamma}{P(D = 1) + 2\gamma - Q} \right]$$

$$\equiv [LB_1, UB_1],$$

(10)

where

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

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

and

$$P(Y = 1|D^* = 0) \in \left[ \frac{P(Y = 1, D = 0) - \delta}{P(D = 0) - 2\delta + Q}, \frac{P(Y = 1, D = 0) + \gamma}{P(D = 0) + 2\gamma - Q} \right]$$

$$\equiv [LB_0, UB_0],$$

(11)

where

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

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

Bounds on the ATE are then given by

$$ATE \in [LB_1 - UB_0, UB_1 - LB_0].$$

However, these bounds are not sharp since it is possible that a different set of measurement error parameters will maximize (minimize) the difference to get
the upper (lower) bound 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:

\[
\begin{align*}
\inf_{b \in [0, \theta_0^{UB} + b]} & \left[ \frac{P[Y = 1, D = 1] - b}{P(D = 1) - b + b} - \frac{P[Y = 1, D = 0] + b}{P(D = 0) + b - b} \right] \\
\leq & \ ATE \\
\leq & \sup_{a \in [0, \theta_0^{UB} - a]} \left[ \frac{P[Y = 1, D = 1] + a}{P(D = 1) + a - a} - \frac{P[Y = 1, D = 0] - a}{P(D = 0) - a + a} \right].
\end{align*}
\]

(12)

Estimation follows by performing separate two-way grid searches for \((b, \tilde{b})\) and \((a, \tilde{a})\) over the feasible region, where:

\[
\begin{align*}
b & \in [0, \min(Q, P[Y = 1, D = 1])], \\
\tilde{b} & \in [0, \min(Q - b, P[Y = 0, D = 0])], \\
a & \in [0, \min(Q, P[Y = 1, D = 0])], \text{ and} \\
\tilde{a} & \in [0, \min(Q - a, P[Y = 0, D = 1])].
\end{align*}
\]

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

If we further impose the assumption of no false positives, equations (8) and (9) simplify to

\[
\begin{align*}
P[Y = 1 | D^* = 1] & = \frac{P(Y = 1, D = 1) + \theta_1^{-}}{P(D = 1) + (\theta_1^{-} + \theta_0^{-})} \quad \text{(13)} \\
P[Y = 1 | D^* = 0] & = \frac{P(Y = 1, D = 0) - \theta_1^{-}}{P(D = 0) - (\theta_1^{-} + \theta_0^{-})}. \quad \text{(14)}
\end{align*}
\]

Bounds on \(P[Y = 1 | D^* = 1]\) are then given by

\[
P[Y = 1 | D^* = 1] \in \left[ \frac{P(Y = 1, D = 1)}{P(D = 1) + \theta_0^{UB^-}}, \frac{P(Y = 1, D = 1) + \theta_1^{UB^-}}{P(D = 1) + \theta_0^{UB^-}} \right] \\
\equiv [LB_1, UB_1]. \quad \text{(15)}
\]

\footnote{The two-way grid search allows the bounds for \(\tilde{b}\) and \(\tilde{a}\) to vary based on the proposed values of \(b\) and \(a\), respectively, as well as \(Q\). Without the inclusion of \(\tilde{b}\) and \(\tilde{a}\), the bounds as presented in Kreider and Pepper (2007) do not necessarily satisfy \(\theta_1^+ + \theta_1^- + \theta_0^+ + \theta_0^- \leq Q\) (Kreider et al. 2013). We are grateful to Brent Kreider for his comments and assistance.}
Similarly, bounds on \( P[Y = 1|D^* = 0] \) are given by

\[
P[Y = 1|D^* = 0] \in \left[ \frac{P[Y = 1, D = 0] - \theta_1^{UB^-}}{P(D = 0) - \theta_0^{UB^-}}, \frac{P[Y = 1, D = 0]}{P(D = 0) - \theta_0^{UB^-}} \right] \\
\equiv [LB_0, UB_0].
\] (16)

Bounds on the ATE are again given by \( ATE \in [LB_1 - UB_0, UB_1 - LB_0] \), which yields

\[
\inf_{\kappa \in [0, \theta_0^{UB^-}]} \left[ \frac{P[Y = 1, D = 1]}{P(D = 1) + h} - \frac{P[Y = 1, D = 0]}{P(D = 0) - h} \right] \leq ATE
\] (17)

\[
\leq \sup_{\alpha \in [0, \theta_1^{UB^-}]} \left[ \frac{P[Y = 1, D = 1] + a}{P(D = 1) + a} - \frac{P[Y = 1, D = 0] - a}{P(D = 0) - a} \right].
\]

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 one.

### 2.3 Worst-Case Selection Bounds

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

\[
P[Y(1) = 1] \in [P[Y = 1, D^* = 1], P[Y = 1, D^* = 1] + P(D^* = 0)]
\]

\[
P[Y(0) = 1] \in [P[Y = 1, D^* = 0], P(D^* = 1) + P[Y = 1, D^* = 0]]
\]

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

Allowing for measurement error, the bounds become

\[
P[Y(1) = 1] \in [P[Y = 1, D = 1] - \theta_1^{UB^-} + \theta_0^{UB^-}, P[Y = 1, D = 1] + P(D = 0) + \theta_0^{UB^-} - \theta_0^-]
\]

\[
P[Y(0) = 1] \in [P[Y = 1, D = 0] + \theta_1^+ - \theta_0^+, P[Y = 1, D = 0] + P(D = 1) - \theta_0^+ + \theta_0^-]
\]

In this case, under the assumption of arbitrary errors,

\[
P[Y(1) = 1] \in [P[Y = 1, D = 1] - \theta_1^{UB^-} + \theta_0^{UB^-}, P[Y = 1, D = 1] + P(D = 0) + \theta_0^{UB^-}]
\] \equiv [LB_1, UB_1]

\[
P[Y(0) = 1] \in [P[Y = 1, D = 0] - \theta_1^{UB^-} - \theta_0^{UB^-}, P[Y = 1, D = 0] + P(D = 1) + \theta_0^{UB^-}]
\] \equiv [LB_0, UB_0],

which yields the following bounds for the ATE:

\[
ATE \in [P[Y = 1, D = 1] - \min\{Q, \theta_1^{UB^-} + \theta_0^{UB^-} \} - P[Y = 1, D = 0] - P(D = 1),

P[Y = 1, D = 1] + P(D = 0) + \min\{Q, \theta_0^{UB^-} + \theta_1^{UB^-} \} - P[Y = 1, D = 0]].
\] (18)
Under the assumption of no false positives,
\[
P[Y(1) = 1] \in [P[Y = 1, D = 1], P[Y = 1, D = 1] + P(D = 0)] \\
\equiv [LB_1, UB_1]
\]
\[
P[Y(0) = 1] \in [P[Y = 1, D = 0] - \theta^UB_1, P[Y = 1, D = 0] + P(D = 1) + \theta^UB_1]
\equiv [LB_0, UB_0],
\]
and the previous bounds simplify to
\[
ATE \in [P[Y = 1, D = 1] - P[Y = 1, D = 0] - P(D = 1) - \theta^UB_1, P[Y = 1, D = 0] + \theta^UB_1].
\]

### 2.4 Monotone Treatment Selection (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, one must be cautious in interpreting each case depending on whether the outcome, $Y$, is desirable or not. Here, we assume that $Y = 1$ corresponds to the desirable outcome.

#### 2.4.1 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. As $Y = 1$ denotes a good outcome, MTSn implies
\[
P[Y(1) = 1|D^* = 0] \geq P[Y(1) = 1|D^* = 1] \\
P[Y(0) = 1|D^* = 0] \geq P[Y(0) = 1|D^* = 1]
\]
We can then rewrite $P[Y(1) = 1]$ as
\[
P[Y(1) = 1] = P[Y(1) = 1|D^* = 1] P(D^* = 1) + P[Y(1) = 1|D^* = 0] P(D^* = 0) \\
= P[Y = 1|D^* = 1 \{1 - P(D^* = 0)\} + P[Y(1) = 1|D^* = 0] P(D^* = 0) \\
= P[Y = 1|D^* = 1] + P(D^* = 0) \{P[Y(1) = 1|D^* = 0] - P[Y = 1|D^* = 1]\},
\]
where the final term is non-negative under MTSn. This implies $P[Y(1) = 1] \geq P[Y = 1|D^* = 1]$. The lower bound for $P[Y(1) = 1]$ is therefore obtained assuming $P[Y(1) = 1|D^* = 0] = P[Y = 1|D^* = 1]$, and the upper bound 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
\[
P[Y(0) = 1] = P[Y(0) = 1|D^* = 1] P(D^* = 1) + P[Y(0) = 1|D^* = 0] P(D^* = 0) \\
= P[Y(0) = 1|D^* = 1] P(D^* = 1) + P[Y(0) = 1|D^* = 0] \{1 - P(D^* = 1)\} \\
= P[Y = 1|D^* = 0] + P(D^* = 1) \{P[Y(0) = 1|D^* = 1] - P[Y(0) = 1|D^* = 0]\},
\]
where the final term is non-positive under MTSn. This implies $P[Y(0) = 1] \leq P[Y = 1|D^* = 0]$. The upper bound is obtained assuming $P[Y(1) = 1|D^* = 0] = P[Y = 1|D^* = 1]$, and the lower bound is obtained assuming $P[Y(0) = 1|D^* = 1] = 0$. This yields

$$P[Y(0) = 1] \leq \left[ P[Y = 1, D^* = 0], \frac{P[Y = 1, D^* = 0]}{P(D^* = 0)} \right].$$

Allowing for measurement error, these bounds become

$$P[Y(1) = 1] \leq \left[ \frac{P(Y(1), D = 1) - \theta_1^+ + \theta_1^-}{P(D = 1)-(\theta_1^+ + \theta_0^+)+ (\theta_1^- + \theta_0^-)}, \frac{P(Y(1), D = 0) + P(Y = 1, D = 1) + \theta_0^+ - \theta_0^-}{P(D = 0) + P(Y = 1, D = 1) + \theta_0^+ - \theta_0^-} \right] = [LB_1, UB_1]$$

$$P[Y(0) = 1] \leq \left[ P[Y = 1, D = 0] + \theta_1^+ - \theta_1^-, \frac{P(Y = 1, D = 0) + \theta_1^+ - \theta_1^-}{P(D = 0) + (\theta_1^+ + \theta_0^+)- (\theta_1^- + \theta_0^-)} \right] = [LB_0, UB_0].$$

Under assumption A1 above (arbitrary errors with upper bound, $P(Z^* = 0) \leq Q$), the lower bound is given in (12) and the upper bound in (18). Similarly, under assumption A2 (no false positives, $P(Z^*|D = 1) = 1$), the lower bound is given in (17) and the upper bound in (19).

### 2.4.2 Positive 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. As $Y = 1$ denotes a good outcome, MTSp implies

$$P[Y(1) = 1|D^* = 1] \geq P[Y(1) = 1|D^* = 0]$$

$$P[Y(0) = 1|D^* = 1] \geq P[Y(0) = 1|D^* = 0].$$

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 non-positive under MTSp. This implies $P[Y(1) = 1] \leq P[Y = 1|D^* = 1]$. The lower bound for $P[Y(1) = 1]$ is obtained assuming $P[Y(1) = 1|D^* = 0] = 0$, and the upper bound is obtained assuming $P[Y(1) = 1|D^* = 0] = P[Y = 1|D^* = 1]$. This yields

$$P[Y(1) = 1] \leq \left[ P[Y = 1, D^* = 1], \frac{P[Y = 1, D^* = 1]}{P(D^* = 1)} \right].$$

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 non-negative under MTSp. This implies $P[Y(0) = 1] \geq P[Y = 1|D^* = 0]$. The lower bound for $P[Y(0) = 1]$ is obtained assuming
\[ P[Y(1) = 1|D^* = 0] = P[Y = 1|D^* = 1], \text{ and the upper bound is obtained assuming } P[Y(0) = 1|D^* = 1] = 1. \text{ This yields} \]

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

Allowing for measurement error, these bounds become

\[ P[Y(1) = 1] \in \left[ P[Y = 1, D = 1] + \theta_1^+ - \theta_1^- + \frac{P[Y = 1, D = 1] + \theta_1^- - \theta_1^+}{P(D = 1) - (\theta_1^+ + \theta_0^-) + (\theta_1^- + \theta_0^+)} \right] \equiv [LB_1, UB_1] \]

\[ P[Y(0) = 1] \in \left[ \frac{P[Y = 1, D = 0] + \theta_1^- - \theta_1^+}{P(D = 0) + (\theta_1^+ + \theta_0^-) - (\theta_1^- + \theta_0^+)} + P[Y = 1, D = 0] + P(D = 1) + \theta_0^- - \theta_0^+ \right] \equiv [LB_0, UB_0]. \]

Under assumption A1 above (arbitrary errors with upper bound, \( P(Z^* = 0) \leq Q \)), the lower bound is given in (18) and the upper bound in (12). Similarly, under assumption A2 (no false positives, \( P(Z^*|D = 1) = 1 \)), the lower bound is given in (19) and the upper bound in (17).

### 2.5 Monotone Treatment Response (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 or not. 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 upper bounds are given by the worst-case upper bounds in (18) and (19) under the assumptions of arbitrary errors and no false positive, respectively. The worst-case lower bounds, however, are now replaced by zero. Combining MTR and MTSn (or MTSp) implies that the upper bounds are given by the upper bounds under MTSn (MTSp), while the lower bounds again are replaced by zero.

### 2.6 Monotone Instrumental Variable (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 with the MTS and MTR assumptions.

#### 2.6.1 MIV and MTS

Let \( \nu \) denote the monotone instrument and assume without loss of generality that \( P[Y(1) = 1] \) and \( P[Y(0) = 1] \) are non-decreasing in \( \nu \). Defining \( u_1 < u < u_2 \)
$u_2$, the MIV assumption implies
\[ P[Y(1) = 1|\nu = u_1] \leq P[Y(1) = 1|\nu = u_2] \]
\[ P[Y(0) = 1|\nu = u_1] \leq P[Y(0) = 1|\nu = u_2]. \]

Recall, the MTSn assumption implies
\[ P[Y(1) = 1|D^* = 1] \leq P[Y(1) = 1|D^* = 0] \]
\[ P[Y(0) = 1|D^* = 1] \leq P[Y(0) = 1|D^* = 0], \]

such that without measurement error the bounds are given by
\[ P[Y(1) = 1] \in \left[ \frac{P[Y = 1, D^* = 1]}{P(D^* = 1)}, P(D^* = 0) + P[Y = 1, D^* = 1] \right] \]
\[ P[Y(0) = 1] \in \left[ P[Y = 1, D^* = 0], \frac{P[Y = 1, D^* = 0]}{P(D^* = 0)} \right]. \]

Combining MIV and MTSn involves the following steps:

1. Split the sample into $J$ cells, $j = 1, 2, ..., J$, based on values of $\nu$, 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 $UB^j_d$ and $LB^j_d$, $j = 1, 2, ..., J$ and $d = 0, 1$.

3. Calculate the overall lower bound for $P[Y(1) = 1]$, denoted by $LB_1$, as
\[ LB_1 = T_n = \sum_j P_j \left( \sup_{j' \leq j} LB^{j'}_1 \right). \]

$T_n$ is therefore a weighted average of the appropriate lower bound estimates across all $J$ cells, constructed by taking $LB_1$ in each $j$th cell and averaging over the supremum of the individual lower bounds across all cells below and including the $j$th cell.

4. Adjust $LB_1$ for finite sample bias by bootstrapping the sampling distribution of $LB_1$.

   (a) Randomly draw with replacement $K$ independent pseudo-samples from the original data of size $N$ with replacement.

   (b) Compute $LB^j_{1k}$, $j = 1, 2, ..., J$ and $k = 1, 2, ..., K$.

   (c) Compute $T^n_k = \sum_j P_{jk} \left( \sup_{j' \leq j} LB^{j'}_{1k} \right)$, $k = 1, 2, ..., K$. This process is identical to that in step (3) above, performed separately for each $k$th pseudo-sample.

---

3Combining MIV with MTSp follows similarly and is therefore excluded for brevity.

4As before, one must be cautious as this assumes $Y = 1$ is a desirable outcome.
(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

\[ \hat{b} = E^*(T_n) - T_n. \]

(f) The bias-corrected LB is given by

\[ LB_1 = T_n^c = T_n - \hat{b} = 2T_n - E^*(T_n). \]

5. Calculate the overall UB for \( P[Y(1) = 1] \), denoted by \( UB_1 \), as

\[ UB_1 = U_n = \sum_j P_j \left( \inf_{j' \geq j} UB_1^{(j')} \right). \]

This assumes \( Y(1) = 1 \) and \( Y(0) = 1 \) are good outcomes. If \( Y(1) = 1 \) and \( Y(0) = 1 \) are bad outcomes, then \( UB_1 = U_n = \sum_j P_j \left( \inf_{j' \leq j} UB_1^{(j')} \right) \). As with the lower bound in step (3), \( T_n \) is again a weighted average of the appropriate upper bound estimates across all \( J \) cells, constructed by taking \( UB_1 \) in each \( j \)th cell and averaging over the infimum of the individual upper bounds across all cells above and including the \( j \)th 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 pseudo-samples from the original data of size \( N \) with replacement.

(b) Compute \( UB_1^{(j')k} \), \( j = 1, 2, \ldots, J \) and \( k = 1, 2, \ldots, K \).

(c) Compute \( U_n^k = \sum_j P_j \left( \inf_{j' \geq j} UB_1^{(j')k} \right) \), \( k = 1, 2, \ldots, K \). This process is identical to that in step (5) above, performed separately for each \( k \)th pseudo-sample.

(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

\[ \hat{b} = E^*(U_n) - U_n. \]

(f) The bias-corrected UB is given by

\[ UB_1 = U_n^c = U_n - \hat{b} = 2U_n - E^*(U_n). \]
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

$$ATE \in [LB_1 - UB_0, UB_1 - LB_0].$$

Prior to continuing, two comments are necessary. First, 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 since 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 utilized in Kreider et al. (2012) is the precision-corrected approach recently proposed in Chernozhukov et al. (2013). This procedure adjusts the terms $LB^j$ and $UB^j$ 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 lower and upper bounds 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 et al. (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.

Second, the asymptotic properties of estimators involving nonsmooth functions like sup and inf, such as those based on a MIV, are the subject of recent debate. Manski and Pepper (2000, p. 1007) note that “the consistency of the resulting bounds estimates is easy to establish.” However, Hirano and Porter (2012, p. 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. As this is the subject of ongoing research (see, e.g., Chernozhukov et al. (2013) and the references therein), users of the tebounds should keep abreast of developments in the literature.

See the author-provided command clrbounds.
2.6.2 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 (i.e., \( P[Y(1) = 1] \) and \( P[Y(0) = 1] \) separately, rather than \( P[Y(1) = 1] - P[Y(0) = 1] \)). As the MTR assumption requires \( P[Y(1) = 1] - P[Y(0) = 1] \geq 0 \), this implies that, within each MIV cell, the lower bound of \( P[Y(1) = 1] \) cannot be strictly less than the upper bound of \( P[Y(0) = 1] \).

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

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

3.2 Specification and Options

The tebounds command requires the user to specify an outcome variable, \( \text{depvar} \), and a treatment assignment variable, \( \text{treat()} \). The remaining options are detailed below:

\( \text{treat(varname)} \) specifies the variable name of the treatment indicator.

\( \text{control(#)} \) specifies the numeric value of \( \text{treat} \) used to indicate the control group, with a default value of 0.

\( \text{treatment(#)} \) specifies the numeric value of \( \text{treat} \) used to indicate the treatment group, with a default value of 1.

\( \text{miv(varname)} \) specifies the variable name of the MIV. The use of \( \text{miv(varname)} \) 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), Manski and Pepper (2009), Kreider et al. (2012), and Lafférs (2013) for a more detailed definition and discussion of proper MIVs.
ncells(#) denotes the number of cells used in obtaining MIV bounds, with a
default value of 5 cells. The MIV variable is divided by percentiles according to
the number of cells specified in ncells. For example, the default of 5 cells will
split the MIV variable into 5 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, with a default string of “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 cor-
rection, with a default of 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 et al.
2013).

bs specifies that confidence intervals be calculated via bootstrap. The level of
the confidence interval is specified in level, which defaults to 95%. Similarly,
the number of replications can be specified in reps(#), with a default of 100
replications.

im specifies that confidence intervals be calculated following Imbens and Manski
(2004). If left unspecified, confidence intervals 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 calcu-
lation as well as the bootstrap replications, if relevant. If bootstrap confidence
intervals 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 re-
sult 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 value of 0 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
pre-specified bootstrap weights available, the weights must be of the form bsw1
bsw2 ... bswN, where N denotes the number of bootstrap replications.⁶

⁶The weights option considers only the non-MIV estimators. The bootstrap weights in
the MIV estimators must be calculated from within the command itself, since these weights
are dependent on the MIV variable as well as the number of cells in ncells.
\texttt{npfu(\#)} specifies the value of \texttt{n(\#)} in the \texttt{bsweights} command. From the \texttt{bsweights} help files, this option specifies how the number of primary sampling units per stratum are handled. The default value of \texttt{-1} indicates a bootstrap sample of size \(n_h - 1\) for all strata, \(h\) (Kolenikov 2010).

\texttt{nodisplay} suppresses the summary results table. Results are still stored in \texttt{ereturn}.

\texttt{graph} specifies that ATE bounds be graphed as a function of the maximum rates of measurement error from \texttt{erates}.

\texttt{saving(string)} specifies a location and stub name for the saving the graphs.

\texttt{replace} indicates that previously saved graphs with the same name should be over-written.

4 Examples

4.1 US School Breakfast Program

Following Millimet and Tchernis (2013), we provide an application of the \texttt{tebounds} command to the study of the U.S. School Breakfast Program (SBP). Specifically, we seek bounds for the average treatment effects of SBP on child weight. The data are from the Early Childhood Longitudinal Study - Kindergarten Class of 1998-99 (ECLS-K) and are available for download from the \textit{Journal of Applied Econometrics} Data Archive.\footnote{http://qed.econ.queensu.ca/jae/datasets/millimet001/} We partially identify the average treatment effect of self-reported SBP participation in first grade (\texttt{break1}) on the probability of being not obese in spring third grade (\texttt{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 (\texttt{ses}) as the MIV, and divide the sample into 20 cells. 95\% bootstrap confidence intervals based on 100 replications are provided. The resulting Stata output follows:

\begin{verbatim}
. tebounds NOTobese, t(break1) erates(0 1 2 5 10) ncells(20) miv(ses) bs > reps(100) gr saving(SJgraphsSBP) replace
\end{verbatim}

\begin{tabular}{l|l|l}
\hline
Outcome: & \texttt{NOTobese} & \texttt{Tebounds} \\
Treatment: & \texttt{break1} & \\
Number of pseudo-samples used in MIV bias correction: & 100 & \\
Number of bootstrap reps for 95\% CIs: & 100 & \\
\hline
\multicolumn{3}{l|}{Error Rate} \\
\multicolumn{3}{l|}{Arbitrary Errors} \\
\multicolumn{3}{l|}{No False Positives} \\
Exogenous Selection Model & & \\
0 & \([-0.059, -0.059]\) p.e. & \([-0.059, -0.059]\) p.e. \\
\hline
\end{tabular}

\footnote{http://qed.econ.queensu.ca/jae/datasets/millimet001/}
<table>
<thead>
<tr>
<th>0</th>
<th>[ -0.076, -0.040] CI</th>
<th>[ -0.076, -0.040] CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>[ -0.095, -0.022] p.e.</td>
<td>[ -0.094, -0.049] p.e.</td>
</tr>
<tr>
<td>.02</td>
<td>[ -0.112, -0.002] CI</td>
<td>[ -0.111, -0.030] CI</td>
</tr>
<tr>
<td>.05</td>
<td>[ -0.146, 0.037] CI</td>
<td>[ -0.146, -0.021] CI</td>
</tr>
<tr>
<td>.1</td>
<td>[ -0.229, 0.144] p.e.</td>
<td>[ -0.229, -0.012] p.e.</td>
</tr>
<tr>
<td></td>
<td>[ -0.246, 0.165] CI</td>
<td>[ -0.245, 0.005] CI</td>
</tr>
<tr>
<td></td>
<td>[ -0.384, 0.267] p.e.</td>
<td>[ -0.384, 0.030] p.e.</td>
</tr>
<tr>
<td></td>
<td>[ -0.399, 0.276] CI</td>
<td>[ -0.399, 0.045] CI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No Monotonicity Assumptions (Worst Case Selection)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.01</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.02</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.05</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Error Rate</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>MTS Assumption: Negative Selection</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.01</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.02</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.05</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>MTS and MTR Assumptions: Negative Selection</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.01</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.02</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.05</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>MIV and MTS Assumptions: Negative Selection</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.01</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.02</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.05</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>MIV, MTS, MTR Assumptions: Negative Selection</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.01</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.02</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.05</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>.1</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

As indicated by the section headings, the output presents the point estimates (p.e.) for the bounds for the ATE, as well as 95% confidence intervals (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 non-participants. However, since 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.

The worst case bounds, assuming no misclassification, have a width of unity and necessarily include zero. 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 lower bound comes from the bounds obtained under exogenous selection, while the upper bound 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 lower bound is replaced by zero. Thus, the bounds still fail to identify the sign of the ATE.

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

Arguably, the MTR assumption may be difficult to justify in the current application. We display this case purely for illustrative purposes.
The final two panels utilize 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 non-obese. 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.

4.2 Monte Carlo

To further illustrate the `tebounds` command, we undertake a simple Monte Carlo exercise. We simulate 10,000 data sets with 2,000 observations from the following data-generating process:
\[ y_0 = I(\nu_0 + 2(z - \epsilon)) \]
\[ y_1 = I(\nu_1 + 2(z - \epsilon)) \]
\[ D^* = I(2*\epsilon + \nu_2) \]
\[ D = \begin{cases} 
1 - D^* & \text{if } U < 0.1 \\
D^* & \text{otherwise} 
\end{cases} \]
\[ \nu_0 \sim N(-1,0.1) \]
\[ \nu_1 \sim N(-2,0.1) \]
\[ \nu_2 \sim N(0,1) \]
\[ \epsilon \sim N(0,1) \]
\[ z \sim N(0,1) \]
\[ U \sim U(0,1) \]

The population ATE is approximately 0.4. \( z \) is a valid MIV as both potential outcomes are increasing in \( z \). The presence of \( \epsilon \) 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 (i.e., 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 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 (i.e., able to exclude zero 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 zero. It is important to note that while the bounds obtained under monotonicity and the assumption of no misclassification exclude zero 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>
<thead>
<tr>
<th>Assumption</th>
<th>Q</th>
<th>Coverage Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exogenous Selection</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.000</td>
</tr>
<tr>
<td>Worst Case</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1.000</td>
</tr>
<tr>
<td>MTSn</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1.000</td>
</tr>
<tr>
<td>MTSn + MTR</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1.000</td>
</tr>
<tr>
<td>MIV + MTSn</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1.000</td>
</tr>
<tr>
<td>MIV + MTSn + MTR</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Q</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exogenous Selection</td>
<td>0</td>
<td>0.890</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.000</td>
</tr>
<tr>
<td>Worst Case</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.000</td>
</tr>
<tr>
<td>MTSn</td>
<td>0</td>
<td>0.890</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.000</td>
</tr>
<tr>
<td>MTSn + MTR</td>
<td>0</td>
<td>0.890</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.000</td>
</tr>
<tr>
<td>MIV + MTSn</td>
<td>0</td>
<td>0.950</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.000</td>
</tr>
<tr>
<td>MIV + MTSn + MTR</td>
<td>0</td>
<td>0.950</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.000</td>
</tr>
</tbody>
</table>

5 Remarks

The `tebounds` command provides a means to partially identify the average treatment effect 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. Confidence intervals are available following Imbens and Manski (2004).
6 References


