Political Analysis (2026), **34**, 23–37
doi:10.1017/pan.2025.10001

ARTICLE

Generalizing Trimming Bounds for Endogenously Missing Outcome Data Using Random Forests

Cyrus Samii[1], Ye Wang[2] and Junlong Aaron Zhou[3]

[1]Department of Politics, New York University, New York, USA; [2]Department of Political Science, University of North Carolina at Chapel Hill, Chapel Hill, USA; [3]Independent Researcher, Mountain View, USA

Corresponding author: Cyrus Samii; Email: cds2083@nyu.edu

(Received 10 April 2024; revised 5 August 2024; accepted 22 November 2024; published online 26 June 2025)

Abstract

We present a method for narrowing nonparametric bounds on treatment effects by adjusting for potentially large numbers of covariates, using generalized random forests. In many experimental or quasi-experimental studies, outcomes of interest are only observed for subjects who select (or are selected) to engage in the activity generating the outcome. Outcome data are thus endogenously missing for units who do not engage, and random or conditionally random treatment assignment before such choices is insufficient to identify treatment effects. Nonparametric partial identification bounds address endogenous missingness without having to make disputable parametric assumptions. Basic bounding approaches often yield bounds that are wide and minimally informative. Our approach can tighten such bounds while permitting agnosticism about the data-generating process and honest inference. A simulation study and replication exercise demonstrate the benefits.

Keywords: causal inference; trimming bounds; partial identification; machine learning; random forest; experiments

Edited by: Jeff Gill

1. Introduction

We present a method for constructing covariate-tightened "trimming bounds" with generalized random forests to address endogenous missingness in causal studies. We first explain trimming bounds (Lee 2009) for those unfamiliar with the method before turning to our contribution: an algorithm for using a potentially large number of covariates to tighten the bounds and generalize an identifying condition (monotonicity).

Causal studies often face endogenously missing outcome data, in which case random or conditionally random treatment assignment is insufficient to point identify causal effects. Below we work with the experiment in Santoro and Broockman (2022), in which subjects were asked to have an online conversation about their "perfect day" with someone from a different political party (an outpartisan). In one group, subjects were informed that their conversation partner was an outpartisan, while in the other group, no such information about their partner was given. Examining only the complete data, the authors found that knowingly talking to an outpartisan was associated with more warmth toward outparty voters in the post-treatment survey. Nevertheless, among the 986 subjects who entered the treatment assignment stage, only 478 completed the conversation and 469 went on to complete the post-treatment survey. The completion rate was higher among those informed about their partner's outpartisan status. For people with a strong interest in politics, a conversation with a "random person" about their perfect day may seem uninteresting while a conversation with an outpartisan might be intriguing. Meanwhile, some people may have a sense of obligation such that they do not drop out if

© The Author(s), 2025. Published by Cambridge University Press on behalf of The Society for Political Methodology.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press

they find the conversation uninteresting. Missingness based on such differences in subjects' types would leave an imbalanced comparison across treatment groups.

Political scientists encounter endogenously missing outcome data regularly. Whether a subject's outcome is revealed often depends on choices. Furthermore, the treatment typically creates asymmetric choice conditions, leading to what Slough (2023) describes as the "phantom counterfactual" problem. In the aforementioned example, knowing the partner's identity may generate two distinct effects: it can reduce the likelihood of completing a conversation while, among those who do complete a conversation, increase the subject's affability toward outpartisans. These are known as the "extensive margin" and the "intensive margin" effects in the literature (Kim, Londregan, and Ratkovic 2019; Staub 2014). If the conversation was not completed, the outcome for the subject is undefined (a "phantom"). Among subjects who completed the conversation, those informed of their partner's partisanship may have a higher degree of political interest as compared to the group completing the conversation in the control group. Consequently, the variation in the average outcome across the two groups is indicative of not only the treatment's impact but also differences attributable to heterogeneity in levels of political interest. As elaborated by Montgomery, Nyhan, and Torres (2018), limiting the analysis to observations with non-missing outcomes implies conditioning on a post-treatment variable (the response indicator), introducing collider biases. If the confounding variables (e.g., political interest) are unmeasured, common imputation or covariate adjustment methods are not justified (Little and Rubin 2019).

One solution is to forgo point identification and rather construct *bounds* on treatment effects (Molinari 2020). The bounds enclose a set of effect values that are consistent with patterns of missingness ("identified set"). The proposed method of "covariate-tightened trimming bounds" generalizes the approach introduced by Lee (2009) to scenarios where a potentially large number of covariates exist and the assumption of monotonic selection (explained below) might be violated. The bounds cover the average causal effect for those who would always have their outcomes observed whether under treatment or control ("always-responders"). In the example of Santoro and Broockman (2022), such an estimand would characterize effects among those who would be willing to engage in an online conversation no matter their treatment status. From a policy perspective, this subgroup defines those for which the intervention raises the fewest red flags in terms of forcing people into an exercise that they find aversive. It is also the subgroup for which intensive margin effects are defined (Slough 2023; Staub 2014).

The logic of trimming bounds is very simple. Suppose those who respond in the control group would always also respond if they had been assigned to the treatment group (monotonic selection). Then the control group consists only of "always-responders." The treatment group is a mixture of always-responders and those who respond only if treated ("treatment-only responders"). The difference in response rates under treatment versus control measures the share of treatment-only responders. We can bound the effect for always-responders using the means of trimmed versions of the treated outcome distribution: a lower bound comes from trimming off the share of treatment-only responders from the top of the distribution, and an upper bound comes from trimming this share from the bottom.

Unfortunately, basic trimming bounds can be very wide. We introduce an approach to tighten them using pre-treatment covariates with the "generalized random forest" (known as *grf*), a machine learning algorithm developed in Wager and Athey (2018) and Athey, Tibshirani, and Wager (2019).[1] Our covariate-tightened trimming bounds are assured to be no wider (in expectation), and can often be substantially narrower, than the basic ones. They are valid when monotonic selection holds only conditional on covariates (as we show in the Santoro and Broockman experiment). We prove that the bounds estimators are consistent and approximately normal in large samples, and we provide approaches for constructing valid confidence intervals. In both simulations and applications, our method works significantly better than the basic trimming bounds and other recent proposals (Olma 2020; Semenova 2025). The method can be generalized to unit-level missingness, binary outcome variables, or the probability of being treated needing to be estimated (as in observational studies). We

[1]Our framework is compatible with other machine learning algorithms. We choose the *grf* for its efficiency of handling potentially large numbers of covariates.

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press

demonstrate with the Santoro and Broockman (2022) experiment, and our Supplementary Material further illustrates performance in the experiment originally studied by Lee (2009), a binary outcome case, and an observational study.

Our covariate-tightened trimming bounds are an alternative for those uncomfortable with either parametric assumptions or untestable exclusion restrictions, as needed in selection modeling methods following Heckman (1979). Our work is in line with other work in political methodology seeking to relax disputable modeling assumptions by using machine learning (Blackwell and Olson 2022; Ratkovic 2022) and partial identification (Duarte *et al.* 2024; Knox, Lowe, and Mummolo 2020). It offers an approach to handling situations in which one cannot avoid conditioning on post-treatment variables like survey response, employment status, survival (Imai 2008; Tchetgen Tchetgen 2014), and endogenous moderators (Blackwell *et al.* 2025).

2. Setup

We consider a randomized or natural experiment with subjects indexed by $i = 1,\ldots,N$, for which we also have measured P pre-treatment covariates collected in the vector $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})$. For each subject, we always observe values of the covariates,[2] the treatment status $D_i \in \{0,1\}$, and the response indicator $S_i \in \{0,1\}$. The realized outcome Y_i is observed only when $S_i = 1$. For $D_i = 1$ and $D_i = 0$, respectively, potential outcomes are given by $(Y_i(1), Y_i(0))$ and potential response indicators are given by $(S_i(1), S_i(0))$. The realized response indicator is $S_i = D_i S_i(1) + (1 - D_i) S_i(0)$, and the realized outcome is $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$ when $S_i = 1$ and missing when $S_i = 0$. Define U_i to be unobserved factors that affect both $(S_i(1), S_i(0))$ and $(Y_i(1), Y_i(0))$.

We make the following assumption on treatment assignment:

Assumption 1. Strong ignorability:

$$(Y_i(1), Y_i(0), S_i(1), S_i(0)) \perp\!\!\!\perp D_i | \mathbf{X}_i,$$
$$\varepsilon < P(D_i = 1 | \mathbf{X}_i) < 1 - \varepsilon, \text{ with } \varepsilon > 0.$$

The first part states that the treatment is conditionally independent to both the potential outcomes and the potential responses. The second part requires that the propensity score, $p(\mathbf{X}_i) = P(D_i = 1 | \mathbf{X}_i)$, is strictly bounded between 0 and 1.[3] The assumption holds in experiments when treatment is randomly or conditionally randomly assigned. In observational studies, the assumption implies that the covariate vector $\mathbf{X}_i$ includes all confounders.

The directed acyclic graph (DAG) in Figure 1 shows relationships that our setting admits. For the Santoro and Broockman (2022) study, U_i could represent the degree of political interest. Our approach also admits the possibility that Y_i directly affects missingness. Only using observations with $S_i = 1$ amounts to conditioning on S_i, which is a "collider" between the treatment D_i and then either the unobserved U_i or the outcome Y_i (Elwert and Winship 2014; Montgomery *et al.* 2018). For example, suppose that for some units, D_i and U_i negatively affect response. For these units, high values of D_i (i.e., being treated) would tend to require low values of U_i for $S_i = 1$ to hold. The lower U_i values would affect the distribution of Y_i values. Correlations induced by conditioning on $S_i = 1$ confound our ability to estimate the causal effect of D_i on Y_i. Such induced correlations result in bias.

To further understand the source of the bias, we classify subjects into four different types based on their responses to the treatment: always-responders ($S_i(0) = S_i(1) = 1$), never-responders ($S_i(0) = S_i(1) = 0$), treatment-only responders ($S_i(0) = 0, S_i(1) = 1$), and control-only responders ($S_i(0) = 1, S_i(1) = 0$). These subgroups are examples of what Frangakis and Rubin (2002) call "principal strata," similar to those in the setting with instrumental variables (Angrist, Imbens, and Rubin 1996).

[2] We discuss unit-level missingness in the Supplementary Material.

[3] We assume that $p(\mathbf{X}_i)$ is known here and discuss estimated propensity scores in the Supplementary Material.

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press

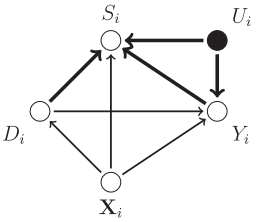


Figure 1. Endogenous missingness.
Note: The above figure shows a directed acyclic graph (DAG) representing an endogenous missingness mechanism. White nodes represent observable variables and black nodes represent unobservable ones. An arrow from one node to the other (a "path") marks the causal relationship from the former to the latter. The heavier arrows indicate the relationships that generate the bias from conditioning on S_i.

Table 1. Principal strata.

Panel 1. Full set of principal strata			Panel 2. Under monotonic selection		
Variable	$S_i = 0$	$S_i = 1$	Variable	$S_i = 0$	$S_i = 1$
$D_i = 0$	NR, TR	CR, AR	$D_i = 0$	NR, TR	AR
$D_i = 1$	NR, CR	TR, AR	$D_i = 1$	NR	TR, AR

Note: AR, NR, TR, and CR in the table stand for always-responders, never-responders, treatment-only responders, and control-only responders, respectively.

These strata are determined by the attributes of the units, such as $\mathbf{X}_i$ and U_i, but not affected by treatment assignment. From Table 1 (Panel 1), we see that among units with observed outcomes ($S_i = 1$), the control group may consist of both control-only responders (CR) and always-responders (AR), while the treatment group is a mixture of treatment-only responders (TR) and always-responders (AR). Therefore, the group-mean difference in the observed outcome does not capture an average of treatment effects for a fixed sub-population.

From the decomposition in Table 1 (Panel 1), we cannot identify the share of units falling into each principal stratum from data. Following Lee (2009), we consider the following assumption on the selection process:

Assumption 2. Monotonic selection:

$$S_i(1) \geq S_i(0)$$

for any *i*.

The assumption requires that under treatment, a unit is at least as likely to respond as under control, thus it excludes the existence of control-only responders. This is shown in Table 1 (Panel 2). The assumption is similar to the "no defiers" assumption for instrumental variables. It should be motivated by arguments about the choice behavior of agents determining response (Slough 2023). We could alternatively assume the opposite: $S_i(1) \leq S_i(0)$ for any *i*. Then, we have control-only responders but not treatment-only responders. Either form of monotonicity allows one to form trimming bounds. In the Santoro and Broockman (2022) example, subjects who would engage when uninformed of their partners' partisanship would be assumed to engage when informed.[4] Parametric and semi-parametric selection models implicitly impose a monotonicity assumption as part of the first-stage regression specification (Vytlacil 2002); the manner in which we state monotonicity here is a nonparametric generalization.

[4]Our reanalysis below suggests that monotonicity does not hold unconditionally in their experiment. We will discuss how to test Assumption 2 and relax it to allow for conditionally monotonic selection in Section 5, following Semenova (2025).

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press

Table 2. Empirical examples

Application	Santoro and Broockman (2022)	Blattman and Annan (2010)	Knox *et al.* (2020)
Y_i	Warmth toward outpartisan voters	Education/distress level	Police use of force in encounter
$D_i = 1$	Knowing partisanship of the partner in conversation	Being kidnapped in conflict	Citizen is minority
$S_i = 1$	Conversation happened and outcome data obtained	Survived and found by researchers	Citizen is under arrest
Mono. selection	Knowing partners' partisanship does not cause subjects to drop out	Avoiding kidnapping does not cause death or disappearance	No encounters in which minority citizens less likely to be arrested
Always-responders	Subjects who will complete conversation regardless of knowledge about partner	Subjects who will survive and be found even if kidnapped	Encounters in which citizen will be arrested regardless of race of citizen
Application	Spilker *et al.* (2018)	Hall, Huff, and Kuriwaki (2019)	Cheema *et al.* (2023)
Y_i	Number of products traded by a firm	Having descendants who fought for the South in the Civil War	Proportion of women who voted in a household
$D_i = 1$	Firm is under a trade agreement	Won the Georgia land lottery and own more slaves	Household was canvassed before election
$S_i = 1$	Firm is an exporter	Having descendants	Household can be tracked in the post-election survey
Mono. selection	Trade agreements do not cause firm not to export	Winning lottery does not cause anyone to have fewer descendants	Being canvassed does not make household less likely to be tracked
Always-responders	Firms that export even without trade agreement	Households that will have descendants regardless of lottery	Households that can always be tracked

Note: This table summarizes how examples from different fields of political science fit into the setup developed here.

Table 2 reports how examples from different fields of political science (American politics, comparative politics, public administration, international relations, and political economy) fit into our setup.

3. Trimming Bounds

Lee (2009) shows that we can use "trimming" to bound the average treatment effect for always-responders, $\tau(1,1) \coloneqq E[Y_i(1) - Y_i(0) | S_i(0) = S_i(1) = 1]$, under Assumptions 1 and 2. The focus on always-responders has substantive motivation. It is the subset of the population that would not withdraw (or be withdrawn) from the intervention under prevailing circumstances. It is also the subset of the

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press

population for which intensive margin effects are defined (as in Lee's original application). Table 2 lists examples.

To lighten the notation in the discussion that follows, suppose that we have unconditionally random assignment. For the more general analysis under strong ignorability (Assumption 1), one would simply switch the expectations to include iterated expectations that condition on and then marginalize over $\mathbf{X}_i$. Given this notational convenience, the average treatment effect for always-responders can be expressed as

$$\begin{aligned}\tau(1,1) &:= E[Y_i(1)|S_i(1) = S_i(0) = 1] - E[Y_i(0)|S_i(1) = S_i(0) = 1] \\ &= E[Y_i|D_i = 1, S_i(1) = S_i(0) = 1] - E[Y_i|D_i = 0, S_i = 1].\end{aligned}$$

The result indicates that under random assignment and monotonic selection, the mean of observed outcomes in the control group identifies the average untreated outcome for the always-responders. Those who respond in the treatment group can belong to either always-responders or treatment-only responders, and so the observed mean does not identify the average treated outcome for the always-responders.

Recall that Assumption 1 implies that the shares of units in each principal stratum are balanced (in expectation) across treatment and control. Under Assumption 2, the rate at which outcomes are observed in the control group, $\Pr[S_i = 1 \mid D_i = 0]$, identifies the share of units in the overall sample that are always-responders. The rate at which outcomes are observed in the treatment group, $\Pr[S_i = 1 \mid D_i = 1]$, identifies the share of units in the overall sample that are either treatment-only responders or always-responders. If we let q denote the share of treated units *with observed outcomes* that are always-responders, then we have

$$q = \frac{\Pr[S_i = 1 \mid D_i = 0]}{\Pr[S_i = 1 \mid D_i = 1]}.$$

The mean of observed outcomes in the treatment group can be written as a mixture of the always-responders mean (with weight q) and the treatment-only responders mean (with weight $1 - q$):

$$\begin{aligned}E[Y_i|D_i = 1, S_i = 1] = {} & qE[Y_i|D_i = 1, S_i(1) = S_i(0) = 1] \\ & + (1-q)E[Y_i|D_i = 1, S_i(1) = 1, S_i(0) = 0].\end{aligned}$$

The quantity $E[Y_i|D_i = 1, S_i(1) = S_i(0) = 1]$ is not identified under random assignment and monotonicity. But we can construct sharp upper and lower bounds that surround it by trimming the lower and upper tails of the observed treated outcome distribution by the share of treatment-only responders, thus retaining the portion with mass q (Lee 2009, Proposition 1a). As shown in Figure 2, in the worst case, always-responders take the bottom (left-most) q share of the treated outcome's distribution. Hence, the average treated outcome of always-responders is bounded from below by $Y(1)$'s expectation over the shaded area in the left plot. Similarly, the average treated outcome for always-responders is bounded from above by $Y(1)$'s expectation over the shaded area in the right plot. Mathematically, we have

$$\begin{aligned}E[Y_i|D_i = 1, S_i = 1, Y_i \leq y_q] &\leq E[Y_i|D_i = 1, S_i(1) = S_i(0) = 1] \\ &\leq E[Y_i|D_i = 1, S_i = 1, Y_i \geq y_{1-q}],\end{aligned}$$

where y_q is the qth-quantile of the observed outcome's conditional distribution in the treatment group that satisfies the condition $\int_{-\infty}^{y_q} dF_{Y|D=1,S=1}(y) = q$. Here, $F_{Y|D=1,S=1}(\cdot)$ is the distribution function for observed treatment group outcomes. The quantile y_{1-q} is similarly defined. We denote $E[Y_i|D_i = 1, S_i = 1, Y_i \leq y_q] - E[Y_i|D_i = 0, S_i = 1]$ as $\tau^L_{TB}(1,1)$ and $E[Y_i|D_i = 1, S_i = 1, Y_i \geq y_q] - E[Y_i|D_i = 0, S_i = 1]$ as $\tau^U_{TB}(1,1)$, where TB is short for trimming bounds. Clearly,

$$\tau^L_{TB}(1,1) \leq \tau(1,1) \leq \tau^U_{TB}(1,1).$$

The interval $[\tau^L_{TB}(1,1), \tau^U_{TB}(1,1)]$ is the identified set for the always-responder treatment effect.

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press

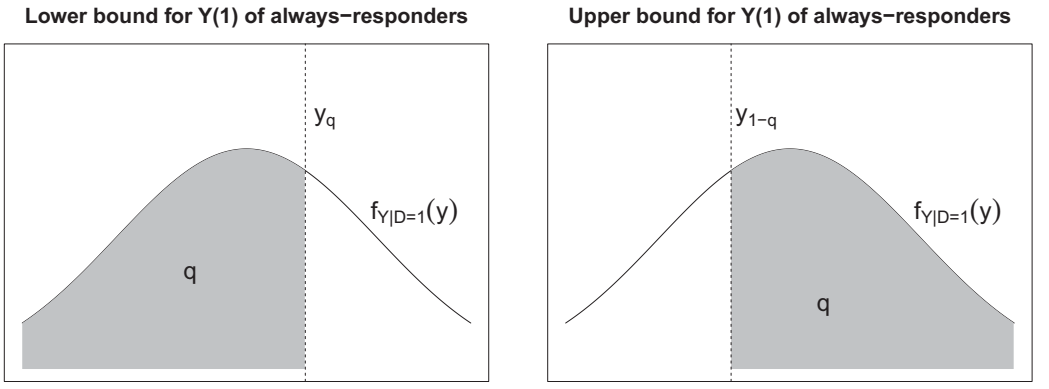


Figure 2. Basic trimming bounds in Lee (2009).

4. Covariate-Tightened Trimming Bounds

Basic trimming bounds that do not incorporate covariate information can be very wide. Lee (2009, Proposition 1b) suggests that when covariates are discrete, researchers can estimate the bounds across blocks formed by them for more precise inference. We extend this approach to cases where the covariates are continuous or large in number. For $\mathbf{X}_i = \mathbf{x}$, consider the always-responders' conditional average treatment effect, $\tau_{\mathbf{x}}(1,1) \coloneqq E[Y_i(1) - Y_i(0) | S_i(1) = S_i(0) = 1, \mathbf{X}_i = \mathbf{x}]$. We can derive the following bounds using the same logic as above:

$$E[Y_i | D_i = 1, S_i = 1, Y_i \le y_{q(\mathbf{x})}(\mathbf{x}), \mathbf{X}_i = \mathbf{x}] \le E[Y_i | D_i = 1, S_i(1) = S_i(0) = 1, \mathbf{X}_i = \mathbf{x}] \le E[Y_i | D_i = 1, S_i = 1, Y_i \ge y_{1-q(\mathbf{x})}(\mathbf{x}), \mathbf{X}_i = \mathbf{x}],$$

where $q(\mathbf{x})$ is the conditional proportion of always-responders among units with observed outcomes in the treatment group; $y_{q(\mathbf{x})}(\mathbf{x})$ and $y_{1-q(\mathbf{x})}(\mathbf{x})$ are the conditional $q(\mathbf{x})$-quantile and $(1-q(\mathbf{x}))$-quantile of the conditional treated outcome distribution.

We denote the conditional control mean for always-responders, $E[Y_i | D_i = 0, S_i = 1, \mathbf{X}_i = \mathbf{x}]$, as $\theta_0(\mathbf{x})$, $E[Y_i | D_i = 1, S_i = 1, \mathbf{X}_i = \mathbf{x}, Y_i \le y_{q(\mathbf{x})}(\mathbf{x})]$ as $\theta_1^L(\mathbf{x})$, and $E[Y_i | D_i = 1, S_i = 1, \mathbf{X}_i = \mathbf{x}, Y_i \ge y_{1-q(\mathbf{x})}(\mathbf{x})]$ as $\theta_1^U(\mathbf{x})$, respectively. Then,

$$\tau^L_{CTB,\mathbf{x}}(1,1) \coloneqq \theta_1^L(\mathbf{x}) - \theta_0(\mathbf{x}) \le \tau_{\mathbf{x}}(1,1) \le \tau^U_{CTB,\mathbf{x}}(1,1) \coloneqq \theta_1^U(\mathbf{x}) - \theta_0(\mathbf{x}),$$

where CTB is short for covariate-tightened trimming bounds. The conditional bounds allow us to see how the bounds vary across different values of the covariates. In practice, we can either estimate the conditional bounds for all the observations or over evaluation points spanning the covariate distribution of the sample. For instance, we can generate the evaluation points by going through the quantiles of one covariate while fixing the values of the other covariates at their sample mode or average.

We can further integrate $(\tau^L_{CTB,\mathbf{x}}(1,1), \tau^U_{CTB,\mathbf{x}}(1,1))$ over the distribution of $\mathbf{X}_i$ among always-responders to recover the unconditional covariate-tightened bounds on $\tau(1,1)$. Define

$$\tau^L_{CTB}(1,1) \coloneqq \frac{\int \tau^L_{CTB,\mathbf{x}}(1,1) q_0(\mathbf{x}) dF(\mathbf{x})}{\int q_0(\mathbf{x}) dF(\mathbf{x})},$$

$$\tau^U_{CTB}(1,1) \coloneqq \frac{\int \tau^U_{CTB,\mathbf{x}}(1,1) q_0(\mathbf{x}) dF(\mathbf{x})}{\int q_0(\mathbf{x}) dF(\mathbf{x})},$$

where

$$q_0(\mathbf{x}) \coloneqq \Pr[S_i(1) = S_i(0) = 1 \mid \mathbf{X}_i = \mathbf{x}] = \Pr[S_i = 1 \mid D_i = 0, \mathbf{X}_i = \mathbf{x}]$$

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press

is the expected conditional proportion of always-takers in the overall sample and $F(\mathbf{x})$ is the distribution of the covariates. The expression indicates that $q_0(\mathbf{x})$ can be identified from the control group. In expectation, covariate-tightened trimming bounds will not be wider than the basic trimming bounds:[5]

$$\tau^L_{TB}(1,1) \leq \tau^L_{CTB}(1,1) \leq \tau(1,1) \leq \tau^U_{CTB}(1,1) \leq \tau^U_{TB}(1,1).$$

The improvement depends on how well the covariates predict the outcome or response rate (Semenova 2025). These bounds are also sharp—we cannot further reduce the identified set's width without additional information.

We can also relax Assumption 2 and allow the direction of monotonicity to change across the values of covariates, as suggested by Semenova (2025). We replace Assumption 2 with the following:

Assumption 3. Conditionally monotonic selection: $S_i(1) \leq S_i(0)|\mathbf{X}_i = \mathbf{x}$ or $S_i(1) \geq S_i(0)|\mathbf{X}_i = \mathbf{x}$.

Under Assumption 3, the covariates space X can be divided into two parts, $\mathcal{X}^+$ and $\mathcal{X}^-$, that satisfy

$$\begin{cases} S_i(1) \leq S_i(0), & \mathbf{X}_i \in \mathcal{X}^- \\ S_i(1) \geq S_i(0), & \mathbf{X}_i \in \mathcal{X}^+ \end{cases} \text{where } \mathcal{X}^+ \cup \mathcal{X}^- = \mathcal{X}.$$

To identify $\mathcal{X}^+$ and $\mathcal{X}^-$, we first estimate the conditional trimming probability, $q(\mathbf{x})$, and approximate $\mathcal{X}^+$ and $\mathcal{X}^-$ with $\hat{\mathcal{X}}^+ \coloneqq \{\mathbf{X}_i : \hat{q}(\mathbf{x}) < 1\}$ and $\hat{\mathcal{X}}^- \coloneqq \{\mathbf{X}_i : \hat{q}(\mathbf{x}) > 1\}$, respectively. Semenova (2025) shows that the classification is accurate in large samples under regularity conditions. If the sample is substantially divided between sets $\hat{\mathcal{X}}^-$ and $\hat{\mathcal{X}}^+$, conditional monotonicity would be more plausible than unconditional monotonicity. Conditional on each part, we can estimate either the conditional or the aggregated bounds. To acquire the aggregated bounds for the whole sample, we take a weighted average the estimates from $\mathcal{X}^+$ and $\mathcal{X}^-$.

Conditionally monotonic selection (Assumption 3) is weaker than its unconditional counterpart (Assumption 2). It also seems uncontroversial that there could be both control-only and treatment-only responders in a sample, even if Lee's original analysis ruled this out. In the Santoro and Broockman (2022) experiment, treatment may induce missingness for those who find it aversive to speak to an outpartisan, while it may reduce missingness for those who find it boring to talk to a "random person." That said, the interpretation of conditionally monotonic selection can be subtle. The assumption holds when, at any point in the covariate space, units for which outcomes are observed are either a mixture of always-responders and control-only responders or a mixture of always-responders and treatment-only responders. If there are values of $\mathbf{X}_i$ in which control-only responders and treatment-only responders are both present, then the assumption is violated. The share of always-responders can vary arbitrarily over covariate values, and so it may not be obvious exactly where in the covariate space transitions occur from regions containing control-only responders versus treatment-only responders. In our algorithm, we use units' covariates to classify them as being in either the (1) always-responders + control-only responders class or (2) always-responders + treatment-only responder class. Our algorithm would be inconsistent if, asymptotically, the covariates are inadequate to perform such a classification. There may be ways to further adjust the bounds to account for such residual classification error. We consider this issue to be beyond the scope of the current paper but one that we think is interesting for further research. We do note that our implementation and accompanying package contains an option to set a tolerance parameter for how sensitive is the classification.

We can also consider the relationship between conditionally monotonic selection and the conditional independence ("missing at random," MAR) assumption that motivates approaches such as imputation, regression adjustment, or inverse propensity score weighting. One can construct scenarios where

[5] Lee (2002, Proposition 1b) proves this when $\mathbf{X}_i$ is independent of D_i, and our Supplementary Material shows this under the more general Assumption 1, allowing for the possibility that basic trimming bounds could be formed under strong ignorability using, for example, inverse propensity score weighting.

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press

conditionally monotonic selection is satisfied while MAR is violated, and vice versa. Importantly, however, the bounding approach allows unmeasured variables to affect both response and the outcome, as demonstrated by our DAG in Figure 1, which we understand to be a first-order threat to identification. In cases where an unobserved variable affects response patterns in a conditional-monotonicity fashion, a MAR approach is biased but we would still be able to recover valid trimming bounds, as we show in Section D.1 of the Supplementary Material. In contrast, when an unobservable variable affects the direction of missingness, even after conditioning on covariates, but is strictly independent of outcomes, then MAR would hold but Assumption 3 would fail. The latter scenario strikes us as more unusual than the former, although we appreciate that this does not exhaust all possibilities.

In Section C of the Supplementary Material, we demonstrate extensions to unit-level missing data, binary outcomes, and estimated propensity scores for generalizing to observational studies under strong ignorability.

5. Estimation

Estimating the covariate-tightened trimming bounds proceeds in two steps. First, we estimate the "nuisance parameters", $(q_0(\mathbf{x}), q_1(\mathbf{x}))$ and $(y_{q(\mathbf{x})}(\mathbf{x}), y_{1-q(\mathbf{x})}(\mathbf{x}))$. Second, we use the nuisance parameter estimates to estimate the "target parameters," that is, the conditional bounds or the aggregated bounds. To estimate parameters evaluated at point $\mathbf{x}$, we use the generalized random forest (*grf*) method. The *grf* method approximates quantities at $\mathbf{x}$ using a weighted average of neighboring observations. To determine the weights, *grf* generates many regression trees from randomly selected sub-samples, and the weight given to observation i for a prediction at $\mathbf{x}$ equals the share of trees in which i falls in the same leaf as the value $\mathbf{x}$. Compared to standard kernel weighting, the *grf* weights are adaptive to the sample distribution and better manage the bias-variance tradeoff. Following Chernozhukov *et al.* (2018), our algorithm also incorporates Neyman orthogonalization and cross-fitting to avoid biases from using the same data to fit nuisance and target parameters, while also maintaining efficiency. The bounds estimators are shown to be asymptotically normal, therefore confidence interval construction is straightforward. Practitioners can also directly construct confidence intervals for $\tau(1,1)$, the ATE for always-responders, using methods developed by Imbens and Manski (2004) and Stoye (2009). Figure 1 summarizes the algorithm. Technical details and statistical properties are given in the Supplementary Material.

Algorithm 1: Honest inference for the covariate-tightened trimming bounds

1 Define the set of covariates $\mathbf{X} = (X_1, \ldots, X_P)$ and evaluation points $\mathcal{X}$ (all sample values for aggregate bounds or specified evaluation points for conditional bounds).
2 Randomly split the sample into K sets $(\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_K)$ with approximately equal size.
3 **for** *each* $k \in \{1, 2, \ldots, K\}$ **do**
4 Fit the $q(\mathbf{X})$ model using the probability forests on $\mathcal{I}_{-k} = \cup_{j \neq k} \mathcal{I}_j$.
5 For any $\mathbf{x} \in \mathcal{X}$, treat $q(\mathbf{x})$ as fixed and fit the models $(y_{q(\mathbf{x})}(\mathbf{X}), y_{1-q(\mathbf{x})}(\mathbf{X}))$ using the quantile forests on $\mathcal{I}_{-k}$.
6 Treat $q(\mathbf{x})$, $y_{q(\mathbf{x})}(\mathbf{x})$ and $y_{1-q(\mathbf{x})}(\mathbf{x})$ as pre-fixed, construct orthogonalized moment conditions, and fit them on $\mathcal{I}_k$.
7 Calculate estimates for the conditional bounds $(\hat{\tau}^U_{CTB,\mathbf{x}}(1,1), \hat{\tau}^L_{CTB,\mathbf{x}}(1,1))$ or aggregated bounds $(\hat{\tau}^U_{CTB}(1,1), \hat{\tau}^L_{CTB}(1,1))$ using the orthogonalized moment conditions.
8 Take the average over the K estimates.
9 Construct confidence intervals using estimated standard errors and critical values from the normal distribution.

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press

6. Simulation

We conduct a Monte Carlo experiment to study the method's performance. We set $N = 1{,}000$ and the number of covariates $P = 10$. Each of the covariates, X_p, is drawn from the uniform distribution on $[0,1]$. The unobservable factor, U, is uniform on $[-1.4, 1.4]$. We only allow X_1 and U to affect both the potential responses $(S_i(0), S_i(1))$ and the potential outcomes $(Y_i(0), Y_i(1))$, while the other nine covariates are pure noise. The setup resembles the reality in which we possess measures of multiple covariates but do not know *ex-ante* which one(s) should be controlled. Our exact data-generating process is as follows:

$$
\begin{aligned}
&Y_i(0) = 1.5 - 2 * U_i^2 + 4 * X_{1i} + \varepsilon_i,\\
&Y_i(1) = Y_i(0) + 2.5 + 1.8 * U_i + 3 * \sin(-0.7 + 2 * X_{1i}),\\
&S_i = \mathbf{1}\{1 - 0.2 * X_{1i} - 1.6 * |U_i| + 0.4 * D_i + 0.1 * D_i * X_{1i} + 2.2 * D_i * |U_i| + \nu_i > 0\},\\
&D_i \sim Bernoulli(0.5), \varepsilon_i \sim \mathcal{N}(0,1), \nu_i \sim \mathcal{N}(0,1).
\end{aligned}
$$

Individualistic treatment effects are nonlinear and non-monotonic in the covariates. Both the selection indicator S_i and outcome Y_i are affected by D_i and U_i, making the sample selection process endogenous. The construction of S_i guarantees that $S_i(1) \geq S_i(0)$ per Assumption 2.

We first estimate the conditional bounds, $(\tau^L_{CTB,\mathbf{x}}(1,1), \tau^U_{CTB,\mathbf{x}}(1,1))$, across 19 points where X_2 to X_{10} are fixed at their sample averages, with X_1 values set at 5ppt. increments. The results are shown in the right plot of Figure 3. Each gray point is an individual-level treatment effect for an observation from one of the Monte Carlo samples. The black line traces out the true CATE for the always-responders; this quantity would be unobservable in a real application and constitutes what the conditional bounds should cover. Upper bounds estimates are depicted as red dots and lower bounds estimates as blue dots. The dotted lines surrounding them are the estimated 95% confidence intervals. The red and blue curves are the true bounds on the always-responders CATEs, based on the simulation parameters. The estimates are very accurate in the middle part of X_1's support but slightly biased at the boundary points.

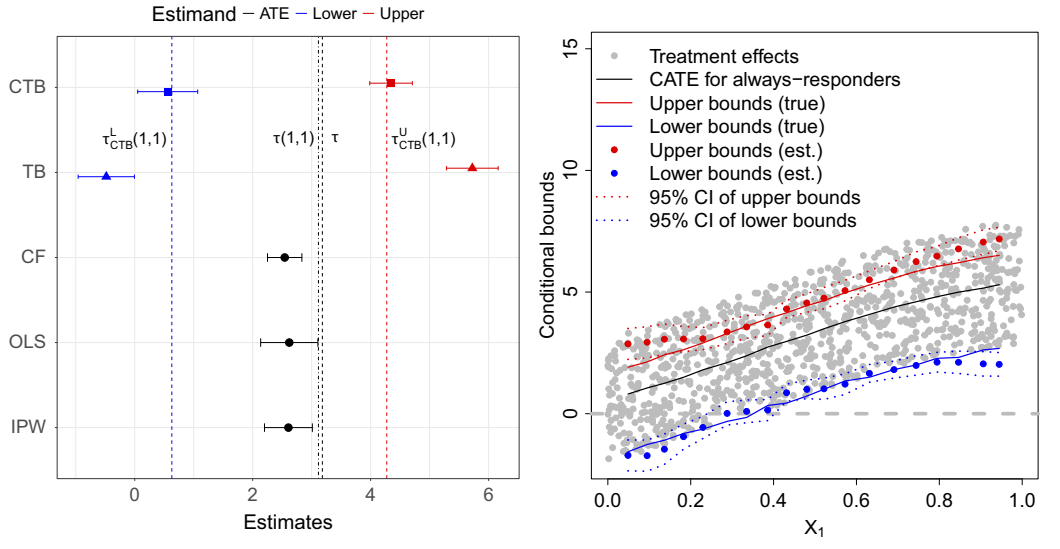


Figure 3. Covariate-tightened trimming bounds in simulated data.
Note: The plot on the left shows the averages of the covariate-tightened trimming bound estimates (CTB, in squares), basic trimming bounds (TB, in triangles), ATE estimate using causal forest on the non-missing sample (CF, in circles), ATE estimate using OLS on the non-missing sample (OLS, in circles), and ATE estimate using OLS re-weighted by the inverse probability of attrition (IPW, in circles). Lower bound estimates are in blue and upper bound estimates are in red. The segments represent 95% confidence intervals. The red and blue dotted lines mark the CTBs. The black dotted line represents the true ATE and the black dash dotted line represents the true ATE for the always-responders. The plot on the right shows the estimates of the conditional covariate-tightened trimming bounds across 19 evaluation points on X_1, holding X_2 to X_{10} to their sample averages.

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press

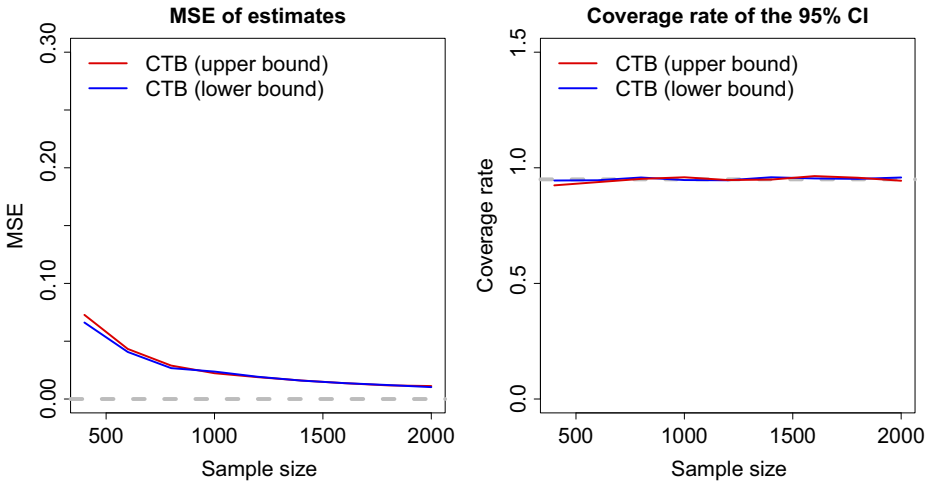


Figure 4. Asymptotic performance of the method.
Note: The left panel shows how the MSE of the lower bound estimate (in blue) and upper bound estimate (in red) varies with sample sizes. The right panel presents the variation of the coverage rate across sample sizes. The gray line marks the nominal level of coverage, 95%.

In the left plot of Figure 3, the top row presents the averages of our CTB estimates with 95% confidence intervals across 1,000 treatment assignments. It is clear that our estimates lead to a narrower identified set of $\tau(1,1)$, the ATE for always-responders, than unadjusted bounds ("TB"). The estimates are also centered around the true CTB and the standard errors of our estimates are not larger than the unadjusted bound estimates. Estimates of the ATE using causal forest or OLS on the non-missing sample and OLS with inverse probability of attrition weighting (IPW) are biased given endogeneity to the unobserved U.[6]

We examine the asymptotic properties of our estimators by varying the sample size from 400 to 2,000. In Figure 4 shows how mean squared error (MSE) and the confidence interval coverage rates perform as N goes from 400 to 2,000, still over 1,000 Monte Carlo draws. MSE declines toward zero as N grows, and our method provides the correct (95%) coverage across the sample sizes. In the Supplementary Material, we present evidence that the performance of the method remains stable when the number of features in the data increases and is superior to alternative methods proposed by Olma (2020) and Semenova (2025).

7. Application

We apply our methods to study the robustness of the findings from Santoro and Broockman (2022) to potentially endogenous attrition. Our Supplementary Material reanalyzes the Job Corps study, which was the original application in Lee (2009), as well as a study by Kalla and Broockman (2022) to demonstrate a binary outcome and a study by Blattman and Annan (2010) to demonstrate an observational study.

We focus on Santoro and Broockman (2022, Study 1), which invited subjects to have a video chat on an online platform with a partner from a different party. The theme of the conversation is what their perfect day would be like. The study started with 986 subjects who satisfied the screening criteria. The subjects were then randomly assigned into either the treatment group ($D_i = 1$), in which they were

[6]We do not compare to outcome imputation, since it works under the same identifying conditions as IPW, or the Heckman correction because we do not have selection instruments that satisfy the exclusion restriction and so identification would depend exclusively on parametric assumptions (Honoré and Hu 2020).

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press

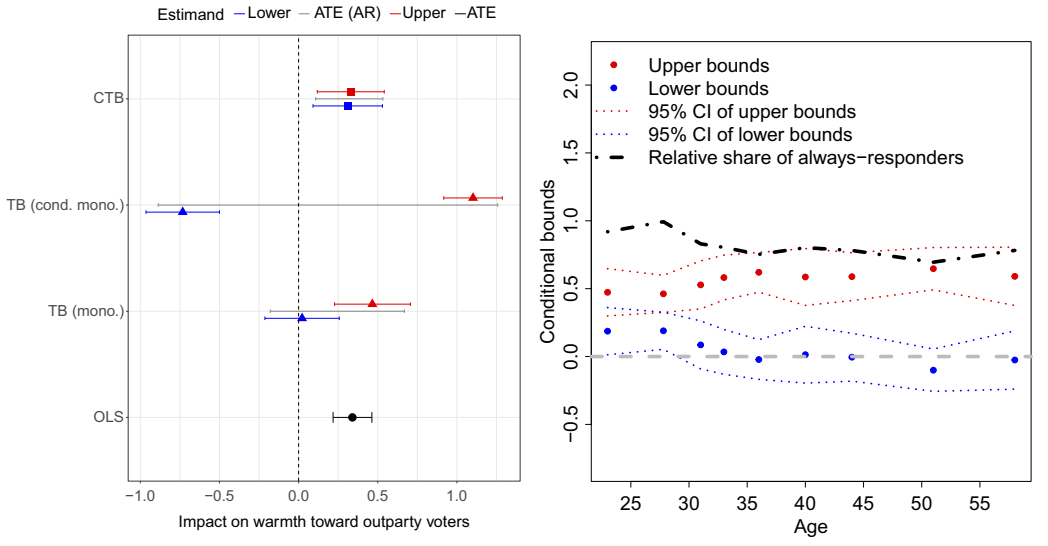


Figure 5. Covariate-tightened trimming bounds in Santoro and Broockman (2022).
Note: In both plots, the outcome variable is warmth toward outpartisan voters. The plot on the left shows the estimated covariate-tightened trimming bounds (CTB, in squares), basic trimming bounds under monotonicity (TB (mono.), in triangles) and under conditional monotonicity (TB (cond. mono.), in triangles), and the ATE estimate using OLS on the non-missing sample (OLS, in circles). Lower bound estimates are in blue and upper bound estimates are in red. Segments represent the 95% confidence intervals for the estimates. Black segments between the red and blue ones represent the 95% confidence intervals for the ATE for always-responders, using the Imbens–Manski approach. The plot on the right shows the estimates of the conditional covariate-tightened trimming bounds across nine evaluation points of age with other demographic attributes are fixed to sample mean or mode.

informed that the partner would be an outpartisan, or the control group ($D_i = 0$), in which they received no extra information. Among subjects that were assigned to a treatment condition, 45.2% of the treated subjects and 39.5% of the control subjects completed the conversation and post-treatment survey. The authors examined the treatment effect on attitudes toward the other parties and found significant effects in the short run.

The authors indicate that the treatment effect on the response rates had a p-value of 0.055, and their omnibus test for covariate balance with respect to education, race/ethnicity, gender, age, and party identification had a p-value of 0.28. Nonetheless, the difference in response rates is not trivial, and other confounding factors may be imbalanced beyond those tested. We can use our covariate-tightened trimming bounds to assess the robustness of their findings. Our inference targets the "always responders" that would complete the conversation and survey regardless of being informed about their partner's partisanship. We focus on the "warmth toward outpartisan voters" outcome (measured by a rescaled thermometer) and rely on the same covariates the authors selected for their analysis, including the age, gender, race, education level, and party identification of the subjects, as well as their pre-treatment outcome.

Figure 5 shows the results. At the bottom is the OLS estimate replicating the main result in the paper. Above that (TB mono.) are the basic trimming bounds, assuming monotonic selection. In this case, however, unconditional monotonicity is questionable. While the overall pattern was such that missingness was lower in the treatment group, some types of people (e.g., those with high affective polarization) may actually choose to drop out *when informed* about their partner's partisan status. When we perform the test proposed in the discussion of conditional monotonicity above, the results suggest that unconditional monotonicity is violated (see the Supplementary Material for details). We thus construct two sets of units ($\mathcal{X}^+$ and $\mathcal{X}^-$) for which treatment is positively and negatively associated with response. Basic trimming bounds (without covariate adjustment) for the first group would be based on bounding the treated mean, and bounds for the second group are based on bounding the control

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press

mean. We can form the basic trimming bounds for the overall sample by constructing the appropriately weighted average of these bounds and the observed treated and control means. The result is shown as "TB (cond. mono.)." We see that relaxing the monotonic selection assumption has strong consequences for our inference. However, when we adjust for covariates using our proposed method ("CTB"), we get tighter results that also allow for conditionally monotonic selection. Both the lower bound (0.272) and the upper bound (0.355) are significantly larger than zero, suggesting that the ATE for the always-responders is indeed positive.

We can examine how the bounds vary over the age of the subjects. We create nine "representative subjects" whose demographic attributes are fixed at the sample mean or mode while their age varies across its deciles. The estimates of the conditional bounds are shown in the right plot of Figure 5. The red and blue dots represent the upper and the lower bounds, respectively, while the dotted lines around them are the 95% confidence intervals. The dash-dotted curve depicts the conditional trimming probability. The response rate (which determines the share of always-responders) is much higher among subjects who are younger than 30. The conditional lower bounds are also significantly larger than zero for this sub-group, while the identified set crosses zero for other subjects. We show results for pre-treatment warmth and ideology in the Supplementary Material.

8. Conclusion

Even in experiments and plausible quasi-experiments, researchers face sample selection due to endogenous nonresponse or censoring. When the selection process is affected by treatment assignment, comparing the average outcome between the treated and untreated subjects no longer yields clean causal estimates, because the two groups consist of different principal strata. Conventional methods attempt to solve the problem by modeling the selection process, using the Heckman correction, inverse probability of attrition weighting, or outcome imputation. But the validity of these methods requires that either conditioning on observables is sufficient or that one knows obscure details of the data-generating process. Inferences are invalid if these conditions are not met. Such approaches do not allow us to be agnostic about the data-generating process and undermine the design-based logic of experimental or quasi-experimental analysis.

We present an alternative partial identification approach that avoids making strong structural assumptions while also improving, in terms of precision, on existing methods. Our proposal combines the trimming bounds method developed by Lee (2009) with a recently developed machine learning algorithm, generalized random forest (Athey *et al.* 2019; Wager and Athey 2018), to generate covariate-adjusted bounds for the average treatment effect on the always-responders. Our approach allows a potentially large set of covariates to tighten the bounds. It also allows one to loosen a "monotonicity" assumption and to still construct informative bounds; we show that the monotonicity assumption is indeed problematic in applied settings. Evidence from our simulation and replication studies shows that the proposed method can be informative when existing methods are not. The price we pay is to forgo a point estimate and rather settle for a range of plausible estimates. In cases where experiments or quasi-experiments are tainted by complex selection problems, we consider this to be a reasonable way to balance credibility and robustness with the need to provide informative estimates (Manski 2019).

Researchers should be aware of some of the trade-offs and limitations of our proposed approach. Using machine learning (the random forest algorithm) offers gains in robustness and efficiency in exchange for the simplicity or interpretability of stratification or linear regression. The algorithm may be harder for non-experts to understand, and one would not obtain results that immediately display partial correlations of covariates with respect to missingness rates, for example. Our algorithm is also more computationally expensive (see Table 1 in the Supplementary Material for running times on a consumer laptop). Finally, our approach requires that there are no points in the covariate space in which both treatment-only and control-only responders reside. Nonetheless, in situations in which the always-responder effect is a quantity of interest and where (conditionally) monotonic selection is plausible, our approach offers a significant improvement over current alternatives.

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press

We have developed an open-source R package, *CTB*, for practitioners to implement the method. A Supplementary Material explains a variety of extensions and applications to experimental and observational studies.

Data Availability Statement. Replication code for this article is available as Wang (2025), and can be accessed via Dataverse at https://doi.org/10.7910/DVN/JDSFJO.

Supplementary Material. For supplementary material accompanying this paper, please visit https://doi.org/10.1017/pan.2025.10001.

Acknowledgements. We thank audiences at ACIC, APSA, MPSA, and POLMETH and the anonymous reviewers, editors, and replication team at *Political Analysis* for helpful feedback.

Author Contributions. All authors contributed to the mathematical analysis, statistical simulations, and empirical statistical analyses.

Conflicts of interest. The authors declare no conflicts of interest.

Ethical Standards. All analyses use publicly available, de-identified replication data.

References

Angrist, J. D., G. W. Imbens, and D. B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–455.

Athey, S., J. Tibshirani, and S. Wager. 2019. "Generalized Random Forests." *The Annals of Statistics* 47 (2): 1148–1178.

Blackwell, M., J. R. Brown, S. Hill, K. Imai, and T. Yamamoto 2025. "Priming Bias versus Post-Treatment Bias in Experimental Designs." *Political Analysis*, forthcoming.

Blackwell, M., and M. P. Olson. 2022. "Reducing Model Misspecification and Bias in the Estimation of Interactions." *Political Analysis* 30 (4): 495–514.

Blattman, C., and J. Annan. 2010. "The Consequences of Child Soldiering." *The Review of Economics and Statistics* 92 (4): 882–898.

Cheema, A., S. Khan, A. Liaqat, and S. K. Mohmand. 2023. "Canvassing the Gatekeepers: A Field Experiment to Increase Women Voters' Turnout in Pakistan." *American Political Science Review* 117 (1): 1–21.

Chernozhukov, V., et al. 2018. "Double/Debiased Machine Learning for Treatment and Structural Parameters." *The Econometrics Journal* 21 (1): C1–C68.

Duarte, G., N. Finkelstein, D. Knox, J. Mummolo, and I. Shpitser 2024. "An Automated Approach to Causal Inference in Discrete Settings." *Journal of the American Statistical Association*, 119 (547): 1778–1793.

Elwert, F., and C. Winship. 2014. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable." *Annual Review of Sociology* 40: 31.

Frangakis, C. E., and D. B. Rubin. 2002. "Principal Stratification in Causal Inference." *Biometrics* 58: 21–29.

Hall, A. B., C. Huff, and S. Kuriwaki. 2019. "Wealth, Slaveownership, and Fighting for the Confederacy: An Empirical Study of the American Civil War." *American Political Science Review* 113 (3): 658–673.

Heckman, J. J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica: Journal of the Econometric Society* 47 (1): 153–161.

Honoré, B. E., and L. Hu. 2020. "Selection without Exclusion." *Econometrica* 88 (3): 1007–1029.

Imai, K. 2008. "Sharp Bounds on the Causal Effects in Randomized Experiments with "Truncation-by-Death"." *Statistics & Probability Letters* 78 (2): 144–149.

Imbens, G. W., and C. F. Manski. 2004. "Confidence Intervals for Partially Identified Parameters." *Econometrica* 72 (6): 1845–1857.

Kalla, J. L., and D. E. Broockman. 2022. ""Outside Lobbying" over the Airwaves: A Randomized Field Experiment on Televised Issue Ads." *American Political Science Review* 116 (3): 1126–1132.

Kim, I. S., J. Londregan, and M. Ratkovic. 2019. "The Effects of Political Institutions on the Extensive and Intensive Margins of Trade." *International Organization* 73 (4): 755–792.

Knox, D., W. Lowe, and J. Mummolo. 2020. "Administrative Records Mask Racially Biased Policing." *American Political Science Review* 114 (3): 619–637.

Lee, D. S. 2002. *Trimming for Bounds on Treatment Effects with Missing Outcomes*. Cambridge: National Bureau of Economic Research.

Lee, D. S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *The Review of Economic Studies* 76 (3): 1071–1102.

Little, R. J., and D. B. Rubin. 2019. *Statistical Analysis with Missing Data*, Vol. 793. Hoboken: John Wiley & Sons.

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press

Manski, C. F. 2019. "Communicating Uncertainty in Policy Analysis." *Proceedings of the National Academy of Sciences* 116 (16): 7634–7641.
Molinari, F. 2020. "Microeconometrics with Partial Identification." *Handbook of Econometrics* 7: 355–486.
Montgomery, J. M., B. Nyhan, and M. Torres. 2018. "How Conditioning on Posttreatment Variables Can Ruin your Experiment and What to Do about it." *American Journal of Political Science* 62 (3): 760–775.
Olma, T. 2020. "Nonparametric Estimation of Truncated Conditional Expectation Functions." *Technical report*, University of Bonn and University of Mannheim, Germany.
Ratkovic, M. 2022. "Relaxing Assumptions, Improving Inference: Integrating Machine Learning and the Linear Regression." *American Political Science Review* 117 (3): 1053–1069.
Santoro, E. and D. E. Broockman 2022. "The Promise and Pitfalls of Cross-Partisan Conversations for Reducing Affective Polarization: Evidence from Randomized Experiments." *Science Advances* 8 (25): eabn5515.
Semenova, V. 2025. "Generalized Lee Bounds." *Journal of Econometrics*, forthcoming.
Slough, T. 2023. "Phantom Counterfactuals." *American Journal of Political Science* 67 (1): 137–153.
Spilker, G., T. Bernauer, I. S. Kim, H. Milner, I. Osgood, and D. Tingley. 2018. "Trade at the Margin: Estimating the Economic Implications of Preferential Trade Agreements." *The Review of International Organizations* 13: 189–242.
Staub, K. E. 2014. "A Causal Interpretation of Extensive and Intensive Margin Effects in Generalized Tobit Models." *Review of Economics and Statistics* 96 (2): 371–375.
Stoye, J. 2009. "More on Confidence Intervals for Partially Identified Parameters." *Econometrica* 77 (4): 1299–1315.
Tchetgen Tchetgen, E. J. 2014. "Identification and Estimation of Survivor Average Causal Effects." *Statistics in Medicine* 33 (21): 3601–3628.
Vytlacil, E. 2002. "Independence, Monotonicity, and Latent Index Models: An Equivalence Result." *Econometrica* 70 (1): 331–341.
Wager, S., and S. Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113 (523): 1228–1242.
Wang, Y. 2025. "Replication Package for "Generalizing Trimming Bounds for Endogenously Missing Outcome Data Using Random Forests" (Samii, Wang, and Zhou, 2024)."

Cite this article: Samii, C., Wang, Y. and Zhou, J. A. (2026). Generalizing Trimming Bounds for Endogenously Missing Outcome Data Using Random Forests. *Political Analysis*, 23–37. https://doi.org/10.1017/pan.2025.10001

https://doi.org/10.1017/pan.2025.10001 Published online by Cambridge University Press