AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

eck for updates

The Stata Journal (2022)
22, Number 2, pp. 344–354 DOI: 10.1177/1536867X221106402

Average treatment effect estimates robust to the "limited overlap" problem: robustate

Yuya Sasaki
Vanderbilt University
Nashville, TN
yuya.sasaki@vanderbilt.edu

Takuya Ura
University of California, Davis
Davis, CA
takura@ucdavis.edu

Abstract. We introduce a new command, `robustate`, that executes the inverse-probability weighting estimation and inference for the average treatment effect with robustness against limited overlap (that is, weak satisfaction of the common support condition). This command produces estimates, standard errors, *p*-values, and confidence intervals for the average treatment effect. The utility of the command is demonstrated with both simulated and real data of right heart catheterization. These illustrations show that the proposed estimator implemented by the `robustate` command indeed exhibits more robustness against limited overlap than the traditional inverse-probability weighting estimator. The main method of the command is proposed in Sasaki and Ura (2022, *Econometric Theory* 38: 66–112).

Keywords: st0674, robustate, average treatment effect, bias correction, common support, inverse-probability weighting, limited overlap, robustness, trimming

1 Introduction

The average treatment effect (ATE) measures the average $E(y_1 - y_0)$ of the difference between the potential outcome under treatment (y_1) and the potential outcome under no treatment (y_0). The difference of the potential outcomes ($y_1 - y_0$) is not directly observed in data, because we observe either y_1 or y_0, but not both, for each individual. Unobserved counterfactual potential outcomes can be inferred from similar observations in terms of covariates or propensity scores. However, such an inference is feasible only when there are similar observations for each observation in a dataset. If the support of the covariates for the treatment group does not overlap with the support of the covariates for the control group, then similar observations in the counterfactual group may not exist for every observation. Thus, the existing methods to estimate the ATE, such as the inverse propensity-score weighting estimator implemented by the `teffects ipw` command and the matching-type estimators implemented by the `teffects psmatch` and `teffects nnmatch` commands, require the common support condition, also known as the strong overlap condition.

Several datasets are known to violate this strong overlap condition and are said to have limited overlaps. Limited overlaps can also be characterized by many observations with propensity-score values close to 0 or 1; for instance, an observation with the propensity-score value of 0.001 is likely untreated with a chance of 99.9%, and you would be unlikely to find a similar observation in the treatment group. For inverse-probability

© 2022 StataCorp LLC

weighting (IPW) estimation, implemented by the `teffects ipw` command, propensity-score values close to 0 or 1 imply infinitesimal denominators and thus outliers. In turn, outliers lead to huge values of standard errors. A common practical solution to this limited overlap problem is to trim observations whose estimated propensity-score values are close to 0 or 1, but this ad hoc procedure generally biases the estimates of the ATE. In Sasaki and Ura (2022), we propose a debiased trimmed IPW estimation procedure, along with a valid standard error of the procedure that accounts for the bias correction. In this article, we introduce the `robustate` command for the debiased trimmed IPW estimation of the ATE, based on our proposed procedure.

We first review the method in sections 2 and 3. The command is introduced in section 4, followed by simulation and real-data analyses in sections 5 and 6, respectively.

2 Review of the method

This section reviews the method of debiased trimmed estimation and inference according to Sasaki and Ura (2022). Specifics and additional practical details implemented by the `robustate` command will be presented in section 3.

IPW estimands take the form of

$$\theta_0 = E\left\{\frac{g_B(\mathbf{w},\boldsymbol{\gamma}_0)}{g_A(\mathbf{w},\boldsymbol{\gamma}_0)}\right\} \tag{1}$$

where $\{g_A(\mathbf{w},\boldsymbol{\gamma}_0), g_B(\mathbf{w},\boldsymbol{\gamma}_0)\}$ is a function of observed random vector $\mathbf{w}$ and an unknown finite-dimensional parameter vector $\boldsymbol{\gamma}_0 \in \Gamma$.[1] The problem of limited overlap is characterized by close-to-0 denominators $g_A(\mathbf{w},\boldsymbol{\gamma}_0)$, which cause large variances in estimating the moment of ratios (1).

A common solution is to trim observations with small denominators $g_A(\mathbf{w},\boldsymbol{\gamma}_0)$. We write an estimator with smooth trimming as

$$\widetilde{\theta}(h_n) = E_n\left[\frac{g_B(\mathbf{w},\widehat{\boldsymbol{\gamma}})}{g_A\left(\mathbf{w},\widehat{\boldsymbol{\gamma}}\right)} \times S\left\{\frac{g_A(\mathbf{w},\widehat{\boldsymbol{\gamma}})}{h_n}\right\}\right] \tag{2}$$

where $\widehat{\boldsymbol{\gamma}}$ is an estimator of $\boldsymbol{\gamma}_0$, h_n is a tuning parameter of trimming, and S is a smoothed indicator function.[2] One idea is to softly trim those observations with small denominator values $g_A(\mathbf{w},\widehat{\boldsymbol{\gamma}})$, where h_n defines “small”. The population counterpart of (2) can be written as

$$\theta(h_n) = E\left[\frac{g_B(\mathbf{w},\boldsymbol{\gamma}_0)}{g_A(\mathbf{w},\boldsymbol{\gamma}_0)} \times S\left\{\frac{g_A(\mathbf{w},\boldsymbol{\gamma}_0)}{h_n}\right\}\right] \tag{3}$$

Estimand (3) with the trimming is different from the original estimand (1), and hence the trimming clearly induces a bias in estimating θ_0 in general.

---

[1] In section 3, we present a specific form of $\{g_A(\mathbf{w},\boldsymbol{\gamma}_0), g_B(\mathbf{w},\boldsymbol{\gamma}_0)\}$ for the case of the ATE.

[2] See section 3 for details about $\widehat{\boldsymbol{\gamma}}$ and S.

Theorem 3 of Sasaki and Ura (2022) motivates that the bias from the trimming can be estimated by

$$\widehat{\lambda}(h_n) = \sum_{\kappa=1}^{k-1} \frac{E_n\left(g_A(\mathbf{w},\widehat{\boldsymbol{\gamma}})^{\kappa-1}\left[S\left\{\frac{g_A(\mathbf{w},\widehat{\boldsymbol{\gamma}})}{h_n}\right\}-1\right]\right)}{\kappa!} \times \widehat{m}^{(\kappa)}(0;\widehat{\boldsymbol{\gamma}}) \tag{4}$$

where

$$\widehat{m}^{(\kappa)}(0;\boldsymbol{\gamma}) = \mathbf{p}_K^{(\kappa)}(0)' E_n\left[\mathbf{p}_K\{g_A(\mathbf{w},\widehat{\boldsymbol{\gamma}})\}\mathbf{p}_K\{g_A(\mathbf{w},\widehat{\boldsymbol{\gamma}})\}'\right]^{-1} E_n\left[\mathbf{p}_K\{g_A(\mathbf{w},\widehat{\boldsymbol{\gamma}})\}g_B(\mathbf{w},\widehat{\boldsymbol{\gamma}})\right]$$

for each $\kappa \in \{1,\ldots,k-1\}$ and $\mathbf{p}_K$ denotes an orthonormal basis of degree K.[3] While this formula for $\widehat{m}^{(\kappa)}(0;\boldsymbol{\gamma})$ may appear complicated, it in fact represents a prediction through the regression of the numerator $g_B(\mathbf{w},\widehat{\boldsymbol{\gamma}})$ on polynomials $\mathbf{p}_K\{g_A(\mathbf{w},\widehat{\boldsymbol{\gamma}})\}$ of the denominator $g_A(\mathbf{w},\widehat{\boldsymbol{\gamma}})$. With the bias estimator (4) subtracted from the biased trimmed estimator (2), we obtain the bias-corrected trimmed estimator

$$\begin{aligned}\widehat{\theta}(h_n) =& E_n\left[\frac{g_B(\mathbf{w},\widehat{\boldsymbol{\gamma}})}{g_A(\mathbf{w},\widehat{\boldsymbol{\gamma}})} \times S\left\{\frac{g_A(\mathbf{w},\widehat{\boldsymbol{\gamma}})}{h_n}\right\}\right] \\ &- \sum_{\kappa=1}^{k-1} \frac{E_n\left(g_A(\mathbf{w},\widehat{\boldsymbol{\gamma}})^{\kappa-1}\left[S\left\{\frac{g_A(\mathbf{w},\widehat{\boldsymbol{\gamma}})}{h_n}\right\}-1\right]\right)}{\kappa!} \times \widehat{m}^{(\kappa)}\left(0;\widehat{\boldsymbol{\gamma}}\right)\end{aligned}$$

After obtaining the debiased trimmed estimator, one must still estimate the variance of this estimator, accounting for the bias estimation as well as an estimation of $\boldsymbol{\gamma}_0$. In Sasaki and Ura (2022), we show in our Lemma 10 that the influence function $\widehat{\theta}(h_n)$ can be written as

$$\begin{aligned} z(h_n) =& \omega_{1,n}(\mathbf{w},\boldsymbol{\gamma}_0) + E\left\{\frac{\partial}{\partial\boldsymbol{\gamma}}\omega_{1,n}(\mathbf{w},\boldsymbol{\gamma})\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}\right\}' \times \boldsymbol{\phi} \\ &+ \sum_{\kappa=1}^{k-1} \frac{\omega_{2,\kappa,n}(\mathbf{w},\boldsymbol{\gamma}_0) \times m^{(\kappa)}(0) + E\{\omega_{2,\kappa,n}(\mathbf{w},\boldsymbol{\gamma}_0)\} \times \psi_\kappa}{\kappa!} \\ &+ \sum_{\kappa=1}^{k-1} \frac{m^{(\kappa)}(0) \times E\left\{\frac{\partial}{\partial\boldsymbol{\gamma}'}\omega_{2,\kappa,n}(\mathbf{w},\boldsymbol{\gamma})\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}\right\} + E\left\{\omega_{2,\kappa,n}(\mathbf{w},\boldsymbol{\gamma}_0)\right\} \times \frac{\partial}{\partial\boldsymbol{\gamma}'} m^{(\kappa)}(0;\boldsymbol{\gamma})\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}}{\kappa!} \times \boldsymbol{\phi} \end{aligned} \tag{5}$$

where ψ_κ and $\boldsymbol{\phi}$ denote the influence functions of $\widehat{m}^{(\kappa)}(0;\boldsymbol{\gamma})$ and $\widetilde{\boldsymbol{\gamma}}$, respectively,[4]

$$\begin{aligned}\omega_{1,n}(\mathbf{w},\boldsymbol{\gamma}) &= g_B(\mathbf{w},\boldsymbol{\gamma})\frac{S\left\{\frac{g_A(\mathbf{w},\boldsymbol{\gamma})}{h_n}\right\}}{g_A(\mathbf{w},\boldsymbol{\gamma})} \\ \omega_{2,\kappa,n}(\mathbf{w},\boldsymbol{\gamma}) &= g_A(\mathbf{w},\boldsymbol{\gamma})^{\kappa-1} \times \left[S\left\{\frac{g_A(\mathbf{w},\boldsymbol{\gamma})}{h_n}\right\} - 1\right]\end{aligned}$$

---

[3] See section 3 for a specific choice and concrete formula of $\mathbf{p}_K$.

[4] See section 3 for specific forms of $\boldsymbol{\phi}$ for the case of logit and probit propensity-score estimators.

In theorem 4 in Sasaki and Ura (2022), we show the asymptotic normality

$$\frac{\widehat{\theta}(h_n) - \theta_0}{\sqrt{\mathrm{Var}\{z(h_n)\}/n}} \xrightarrow{d} \mathcal{N}(0,1)$$

under suitable assumptions.

The above theory motivates $\widehat{\theta}(h_n)$ as a debiased trimmed estimator of θ_0 and

$$\sqrt{\widehat{\mathrm{Var}}\{z(h_n)\}/n}$$

as its asymptotically valid standard error, accounting for the bias correction and a preliminary estimation of $\boldsymbol{\gamma}_0$. The `robustate` command uses these expressions to produce its output, which consists of estimates, standard errors, p-values, and confidence intervals for the ATE.

3 Specifics and additional details about the method

While section 2 presents a review of the general method, this section focuses on the ATE and presents specifics and additional details about the method implemented by the `robustate` command.

3.1 ATE

Suppose that a researcher observes a random sample of $\mathbf{w} = (y, d, \mathbf{x}')'$, where y denotes an observed outcome, d denotes the binary indicator of an observed treatment, and $\mathbf{x}$ denotes the vector of observed controls. Suppose that $p(\mathbf{x}, \boldsymbol{\gamma}_0)$ models the conditional probability $P(d = 1|\mathbf{x})$ of receiving the treatment given the observed characteristics $\mathbf{x}$, also known as the propensity score.[5] With these notations, the IPW estimator of the ATE can be written by (1), where $\{g_A(\mathbf{w}, \boldsymbol{\gamma}_0), g_B(\mathbf{w}, \boldsymbol{\gamma}_0)\}$ is defined by

$$\begin{aligned} g_A(\mathbf{w}, \boldsymbol{\gamma}_0) &= d + \{P(\mathbf{x}, \boldsymbol{\gamma}_0) - 1\} \times (2d - 1) \\ g_B(\mathbf{w}, \boldsymbol{\gamma}_0) &= y \times (2d - 1) \end{aligned}$$

There are alternative ways to define $\{g_A(\mathbf{w}, \boldsymbol{\gamma}_0), g_B(\mathbf{w}, \boldsymbol{\gamma}_0)\}$ for the ATE. We use this particular definition of $\{g_A(\mathbf{w}, \boldsymbol{\gamma}_0), g_B(\mathbf{w}, \boldsymbol{\gamma}_0)\}$ to conveniently trim observations with propensity scores close to both 0 and 1 by (2) [note that $g_A(\mathbf{w}, \boldsymbol{\gamma}_0) = 1 - P(\mathbf{x}, \boldsymbol{\gamma}_0)$ if $d = 0$ and $g_A(\mathbf{w}, \boldsymbol{\gamma}_0) = P(\mathbf{x}, \boldsymbol{\gamma}_0)$ if $d = 1$].

---

[5] See section 3.2 for more details.

3.2 Propensity-score estimation

Suppose that the propensity-score function is modeled by the generalized linear model $p(\mathbf{x}, \boldsymbol{\gamma}_0) = \Lambda(\mathbf{x}'\boldsymbol{\gamma}_0)$ for some link function Λ. The `robustate` command contains two options of Λ. One option is the standard logistic cumulative distribution function (CDF),

$$\Lambda(z) = 1/\{1 + \exp(-z)\}$$

in which case $p(\mathbf{x}, \boldsymbol{\gamma}_0) = \Lambda(\mathbf{x}'\boldsymbol{\gamma}_0)$ represents the logit binary treatment choice model. The other option is the standard normal CDF,

$$\Lambda(z) = \Phi(z)$$

in which case $p(\mathbf{x}, \boldsymbol{\gamma}_0) = \Lambda(\mathbf{x}'\boldsymbol{\gamma}_0)$ represents the probit binary treatment choice model.

The parameter vector $\boldsymbol{\gamma}_0$ of the generalized linear model is estimated by the maximum-likelihood estimation procedure. Concretely, the estimator $\widetilde{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}_0$ is defined by

$$\widetilde{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma} \in \Gamma} Q_n(\boldsymbol{\gamma}) \tag{6}$$

where the criterion function $Q_n(\boldsymbol{\gamma})$ takes the form of

$$Q_n(\boldsymbol{\gamma}) = E_n \left[d \log \{\Lambda(\mathbf{x}'\boldsymbol{\gamma})\} + (1-d) \log \{1 - \Lambda(\mathbf{x}'\boldsymbol{\gamma})\} \right]$$

3.3 Influence function of preliminary estimation

The asymptotic variance of the debiased trimmed estimator (2) needs to account for the preliminary estimation of $\boldsymbol{\gamma}_0$. As such, the influence function representation (5) of $\widehat{\theta}(h_n)$ involves the influence function $\boldsymbol{\phi}$ of the preliminary estimator $\widetilde{\boldsymbol{\gamma}}$. In light of the concrete estimator $\widetilde{\boldsymbol{\gamma}}$ defined in (6), the influence function $\boldsymbol{\phi}$ of the preliminary estimator $\widetilde{\boldsymbol{\gamma}}$ can be estimated by

$$\widehat{\boldsymbol{\phi}} = -\{d^2 Q_n(\widetilde{\boldsymbol{\gamma}})\}^{-1} \frac{d}{d\boldsymbol{\gamma}} \left[d \log \{\Lambda(\mathbf{x}'\boldsymbol{\gamma})\} + (1-d) \log \{1 - \Lambda(\mathbf{x}'\boldsymbol{\gamma})\} \right] \Big|_{\boldsymbol{\gamma}=\widetilde{\boldsymbol{\gamma}}} \tag{7}$$

In the special case of the logit link function, it can be more explicitly written with the analytic expression

$$\widehat{\boldsymbol{\phi}} = E_n \left\{ \mathbf{x} \Lambda(\mathbf{x}'\widetilde{\boldsymbol{\gamma}}) \mathbf{x}' \right\}^{-1} \mathbf{x} \{d - \Lambda(\mathbf{x}'\widetilde{\boldsymbol{\gamma}})\}$$

The `robustate` command uses this explicit formula for the case of the logit propensity-score estimation, while it uses numerical derivatives to (7) for the case of the probit propensity-score estimation.

3.4 Smoothed indicator function

The smoothed indicator function S can be defined arbitrarily as long as it satisfies $S(0) = 0$, $S(u) = 1$ for $u \geq 1$, and some regularity conditions. In the implementation by the `robustate` command, the specific form of S defined by

$$S(u) = \begin{cases} 0 & \text{if } u < 0 \\ 6u^5 - 15u^4 + 10u^3 & \text{if } 0 \leq u \leq 1 \\ 1 & \text{if } 1 < u \end{cases}$$

is used. With this definition of S, the smooth trimmed estimator (2) uses the full unit weight for those observations with a large denominator $[g_A(\mathbf{w}, \widehat{\gamma}) \geq h_n]$ and a smaller subunit weight for those observations with a small denominator $[g_A(\mathbf{w}, \widehat{\gamma}) < h_n]$. The relatively high degrees (five degrees) of polynomial are needed to satisfy the regularity conditions in theorem 5 of Sasaki and Ura (2022).

3.5 Shifted orthonormal Legendre polynomial basis

The bias correction in (4) is based on a sieve nonparametric estimation with an orthonormal basis $\mathbf{p}_K(a)$ of degree K. The `robustate` command uses the shifted orthonormal Legendre polynomial basis, which is given by

$$\mathbf{p}_K(a) = \left\{ \begin{array}{c} 1 \\ \sqrt{3}(2a-1) \\ \sqrt{5}(6a^2 - 6a + 1) \\ \sqrt{7}(20a^3 - 30a^2 + 12a - 1) \\ \sqrt{9}(70a^4 - 140a^3 + 90a^2 - 20a + 1) \\ \sqrt{11}(252a^5 - 630a^4 + 560a^3 - 210a^2 + 30a - 1) \\ \vdots \end{array} \right\}$$

The minimum value of K that is allowed in theory is 4. The default value of K in the `robustate` command is set to 4 (see section 2.2).

4 The robustate command

4.1 Syntax

The syntax of the `robustate` command is as follows:

`robustate` *outcome treatment controls* [*if*] [*in*] [`, probit h(`*real*`) k(`*real*`)`]

outcome stands for the outcome variable y, *treatment* stands for the binary indicator of a treatment d, and *controls* include observed controls $\mathbf{x}$. Exactly one *outcome* variable, exactly one *treatment* variable, and at least one *controls* variable must be included to run the command.

4.2 Options

`probit` specifies to estimate the propensity score with probit estimation. The default is to use logit estimation.

`h(`*real*`)` sets the trimming threshold. The default is `h(0.1)`. *real* must be a real number in $(0, 1)$. Larger values induce larger biases of the naïve estimator.

`k(`*real*`)` sets the sieve dimension for bias correction. The default is `k(4)`. *real* must be an integer no smaller than 4.

4.3 Stored results

`robustate` stores the following in `e()`:

Scalars	
`e(N)`	number of observations
`e(h)`	trimming threshold
`e(k)`	order of orthonormal basis
Macros	
`e(cmd)`	`robustate`
`e(properties)`	`b V`
`e(pscore)`	`logit` or `probit`
Matrices	
`e(b)`	coefficient vector
`e(V)`	variance–covariance matrix of the estimators
Function	
`e(sample)`	marks estimation sample

5 Simulation studies

In this section, we use Monte Carlo simulations to evaluate the small-sample performance of the method implemented by the `robustate` command. We use the generating design outlined below.

First, generate the five-dimensional vector $\mathbf{x} = (x_1, \ldots, x_5)'$ of controls according to the t distribution of df degrees of freedom:

$$x_1, \ldots, x_5 \overset{\text{i.i.d.}}{\sim} t(\text{df})$$

In turn, the propensity score p is produced by

$$p = \Lambda(\mathbf{x}'\boldsymbol{\gamma}_0)$$

where Λ is the logistic CDF and $\boldsymbol{\gamma}_0 = (1.0, 0.8, 0.6, 0.4, 0.2)'$. Note that a smaller value of df induces a heavier-tailed distribution of $x_1, \ldots, x_5$, which in turn causes the propensity score p to be closer to 0 and 1. In other words, the smaller df is, the more intensive limited overlaps become.

Given the treatment choice d and the propensity score p, the outcome variable is finally generated by

$$\begin{aligned} y &= (1-d) \times y_0 + d \times y_1 \\ y_0 &= v_0 \\ y_1 &= v_1 + \left\{1 + 0.5 \times \Phi^{-1}(p)\right\} \end{aligned}$$

where $v_0, v_1 \overset{\text{i.i.d.}}{\sim} N(0,1)$ and Φ^{-1} denotes the quantile function of the standard normal distribution. Recall that y_0 and y_1 denote the potential outcomes under treatment and no treatment, respectively, which are not usually observed by a researcher. Note that the inclusion of p in the equation of the potential outcome y_1 plays the role of causing an endogeneity.

We vary the sample size $n \in \{200, 400\}$ and the degrees-of-freedom parameter df $\in \{3, 5, 7, 9\}$ across sets of simulations. Each set of simulations is based on 500 Monte Carlo iterations. We summarize the simulation results in the table below. For each of the naïve IPW estimators and our proposed robust IPW estimator, the table reports the root mean squared error (RMSE), the 90% coverage frequency, and the 95% coverage frequency.

		Naïve IPW			Robust IPW		
n	df	RMSE	90%	95%	RMSE	90%	95%
200	9	**0.220**	0.922	0.964	**0.179**	0.890	0.944
400	9	**0.148**	0.916	0.962	**0.126**	0.912	0.958
200	7	**0.217**	0.920	0.962	**0.180**	0.896	0.958
400	7	**0.156**	0.906	0.966	**0.129**	0.906	0.956
200	5	**0.244**	0.912	0.968	**0.187**	0.900	0.948
400	5	**0.224**	0.892	0.954	**0.126**	0.910	0.954
200	3	**0.831**	0.890	0.948	**0.194**	0.884	0.934
400	3	**0.346**	0.896	0.954	**0.134**	0.884	0.944

Observe that the coverage frequencies are fairly close to the respective nominal probabilities for both estimators. On the other hand, the RMSE is strikingly different between the naïve and robust estimators. Specifically, under each sample size n and each degree of freedom df, the RMSE of the naïve estimator is much larger than that of the robust estimator. Furthermore, this difference tends to become bigger as the df of the t distribution decreases. Recall that the problem of limited overlap is more intensive as df gets smaller. Hence, these results show that the naïve estimator is vulnerable to limited overlap, while our robust estimator is literally robust against it.

6 Illustration of the command

We now illustrate the `robustate` command with an analysis of the treatment effect of catheterization on 30-day survival using a subsample of real data. A small-sample dataset is available in the `robustate` package, and it can be loaded by typing the following command:

```
. use catheterization_small
```

The command below executes the estimation and inference for the ATE of catheterization on 30-day survival.

```
. robustate outcome treat age alb1 amihx aps1 bili1 ca_meta ca_yes card cardiohx
> cat1_chf cat1_cirr cat1_colon cat1_coma cat1_copd cat1_lung cat1_mosfmal
> cat1_mosfsep cat2_cirr cat2_colon cat2_coma cat2_lung cat2_mosfmal
> cat2_mosfsep chfhx chrpulhx crea1 das2d3pc dementhx
```

Here `outcome` is the name of the outcome variable, `treat` is the name of the binary indicator of a treatment, and all the rest are the names of control variables.

This command produces the results displayed below. The first row of this output shows an estimate based on the naïve inverse propensity-score weighting estimation method, which is not robust against limited overlap. The second row, on the other hand, shows an estimate based on our proposed debiased trimmed estimation method, which is robust against limited overlap.

```
                                                   Observations:         5735
                                                   P-Score Estimation:  Logit
Average Treatment Effect (ATE):
       Naive Inverse Propensity Score-Weighed Estimation (naiveATE) &
       Robust Inverse Propensity Score-Weighed Estimation (robustATE)
------------------------------------------------------------------------------
             |  Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
    naiveATE |    -.053216   .0432941    -1.23   0.219    -.1380709     .031639
   robustATE |   -.1043876   .0311707    -3.35   0.001    -.1654811   -.0432941
------------------------------------------------------------------------------

  (output omitted)
```

Notice that the robust estimator yields a much narrower 95% confidence interval than the naïve estimator. These results are consistent with the simulation results presented in section 5. It is also worth noting that the robust method indicates a statistically significant ATE, whereas the naïve method does not. The counterintuitive conclusion that the right heart catheterization has negative effects on 30-day survivals is consistent with the previous finding.

7 Conclusion

In this article, we introduced the `robustate` command, which executes an IPW estimation and inference for the ATE with robustness against limited overlap based on the method of Sasaki and Ura (2022). We illustrated the command with simulated data and demonstrated that the proposed method achieves smaller RMSEs than the traditional estimator, especially as the intensity of the limited overlap becomes more severe. We also illustrated the command with a small real dataset of right heart catheterization and showed that the proposed method delivers a narrower confidence interval than the traditional estimator.

While a brief review of the main method of the debiased trimmed IPW estimation is presented in section 2 and additional practical details of the method implemented by the `robustate` command are described in section 3, we refer interested readers to Sasaki and Ura (2022) for further details about the supporting econometric theory.

Finally, we close this article by discussing how other approaches might treat the limited overlap problem. A common approach is to trim observations with estimated propensity scores close to 0 or 1 (where popular trimming thresholds are 0.1 and 0.9) and to run the standard IPW estimator without bias correction as implemented by the `teffects ipw` command. This approach biases the estimate with respect to the ATE in general but instead obtains the conditional ATE among the subpopulation of the untrimmed region (for example, $[0.1, 0.9]$) of the propensity score (see Crump et al. [2009]). Furthermore, a naïve application of the existing commands, like `teffects ipw`, would not obtain a correct standard error even for such a conditional ATE, because of the data-driven trimming. Matching-type estimators, as implemented by the `teffects psmatch` command and the `teffects nnmatch` command, would generally bias the estimate with respect to the ATE because matching observations outside of the common support entails matching dissimilar observations. We are not aware of a bias-correction method for matching-type estimators.

8 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 22-2
. net install st0674     (to install program files, if available)
. net get st0674         (to install ancillary files, if available)
```

`robustate` can also be downloaded from the Statistical Software Components Archive by typing

```
. ssc install robustate
```

9 References

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik. 2009. Dealing with limited overlap in estimation of average treatment effects. *Biometrika* 96: 187–199. https://doi.org/10.1093/biomet/asn055.

Sasaki, Y., and T. Ura. 2022. Estimation and inference for moments of ratios with robustness against large trimming bias. *Econometric Theory* 38: 66–112. https://doi.org/10.1017/S0266466621000025.

About the authors

Yuya Sasaki is an associate professor of economics at Vanderbilt University.

Takuya Ura is an assistant professor of economics at University of California, Davis.