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

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

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

http://www.stata-journal.com

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

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$ 98	1-year subscription	$138
2-year subscription	$165	2-year subscription	$245
3-year subscription	$225	3-year subscription	$345
1-year student subscription	$ 75	1-year student subscription	$ 99
1-year institutional subscription	$245	1-year institutional subscription	$285
2-year institutional subscription	$445	2-year institutional subscription	$525
3-year institutional subscription	$645	3-year institutional subscription	$765
Electronic only		**Electronic only**	
1-year subscription	$ 75	1-year subscription	$ 75
2-year subscription	$125	2-year subscription	$125
3-year subscription	$165	3-year subscription	$165
1-year student subscription	$ 45	1-year student subscription	$ 45

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

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

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

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

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

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

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

The bmte command: Methods for the estimation of treatment effects when exclusion restrictions are unavailable

Ian McCarthy
Emory University
Atlanta, GA
ianmccarthy.econ@gmail.com

Daniel Millimet
Southern Methodist University
Dallas, TX
Institute for the Study of Labor
Bonn, Germany
millimet@smu.edu

Rusty Tchernis
Georgia State University
Atlanta, GA
Institute for the Study of Labor
Bonn, Germany
National Bureau of Economic Research
Cambridge, MA
rtchernis@gsu.edu

Abstract. We present a new Stata command, `bmte` (bias-minimizing treatment effects), that implements two new estimators proposed in Millimet and Tchernis (2013, *Journal of Applied Econometrics* 28: 982–1017) and designed to estimate the effect of treatment when selection on unobserved variables exists and appropriate exclusion restrictions are unavailable. In addition, the `bmte` command estimates treatment effects from several alternative estimators that also do not rely on exclusion restrictions for identification of the causal effects of the treatment, including the following: 1) Heckman's two-step estimator (1976, *Annals of Economic and Social Measurement* 5: 475–492; 1979, *Econometrica* 47: 153–161); 2) a control function approach outlined in Heckman, LaLonde, and Smith (1999, *Handbook of Labor Economics* 3: 1865–2097) and Navarro (2008, *The New Palgrave Dictionary of Economics* [Palgrave Macmillan]); and 3) a more recent estimator proposed by Klein and Vella (2009, *Journal of Applied Econometrics* 24: 735–762) that exploits heteroskedasticity for identification. By implementing two new estimators alongside preexisting estimators, the `bmte` command provides a picture of the average causal effects of the treatment across a variety of assumptions. We present an example application of the command following Millimet and Tchernis (2013, *Journal of Applied Econometrics* 28: 982–1017).

Keywords: st0355, bmte, treatment effects, propensity score, unconfoundedness, selection on unobserved variables

1 Introduction

The causal effect of binary treatment on outcomes is a central component of empirical research in economics and many other disciplines. When individuals self-select into treatment and when prospective randomization of the treatment and control groups is not feasible, researchers must adopt alternative empirical methods intended to control for the inherent self-selection. If individuals self-select on the basis of observed variables (selection on observed variables), a variety of appropriate methodologies are available to estimate the causal effects of the treatment. If instead individuals self-select on the basis of unobserved variables (selection on unobserved variables), estimating treatment effects is more difficult.

When one is confronted with selection on unobserved variables, the most common empirical approach is to rely on an instrumental variable (IV); however, if credible instruments are unavailable, a few approaches now exist that attempt to estimate the effects of the treatment without an exclusion restriction. This article introduces a new Stata command, `bmte`, that implements two recent estimators proposed in Millimet and Tchernis (2013) and designed to estimate treatment effects when selection on unobserved variables exists and appropriate exclusion restrictions are unavailable:

i. The minimum-biased (MB) estimator: This estimator searches for the observations with minimized bias in the treatment-effects estimate of interest. This is accomplished by trimming the estimation sample to include only observations with a propensity score within a certain interval as specified by the user. When the conditional independence assumption (CIA) holds (that is, independence between treatment assignment and potential outcomes, conditional on observed variables), the MB estimator is unbiased. Otherwise, the MB estimator tends to minimize the bias among estimators that rely on the CIA. Furthermore, the MB estimator changes the parameter being estimated because of the restricted estimation sample.

ii. The bias-corrected (BC) estimator: This estimator relies on the two-step estimator of Heckman's bivariate normal (BVN) selection model to estimate the bias among estimators that inappropriately apply the CIA (Heckman 1976, 1979). However, unlike the BVN estimator, the BC estimator does not require specification of the functional form for the outcome of interest in the final step. Moreover, unlike the MB estimator, the BC estimator does not change the parameter being estimated.

In addition, the `bmte` command summarizes results of several alternative estimators across a range of assumptions, including standard ordinary least-squares (OLS) and inverse-probability-weighted (IPW) treatment-effects estimates. The `bmte` command also presents the results of additional estimates applicable when the CIA fails and valid exclusion restrictions are unavailable, including the following: 1) Heckman's BVN estimator; 2) a control function (CF) approach outlined in Heckman, LaLonde, and Smith (1999) and Navarro (2008); and 3) a more recent estimator proposed by Klein and Vella (2009) that exploits heteroskedasticity for identification. By implementing two new es-

timators alongside preexisting estimators, the `bmte` command provides a picture of the average causal effects of the treatment across a variety of assumptions and when valid exclusion restrictions are unavailable.

2 Framework and methodology

Here we provide a brief background on the potential-outcomes model and the estimators implemented by the `bmte` command. For additional discussion, see Millimet and Tchernis (2013). We consider the standard potential-outcomes framework, denoting by $Y_i(T)$ the potential outcome of individual i under binary treatment $T \in \mathcal{T} = (0, 1)$. The causal effect of the treatment ($T = 1$) relative to the control ($T = 0$) is defined as the difference between the corresponding potential outcomes, $\tau_i = Y_i(1) - Y_i(0)$.

In the evaluation literature, several population parameters are of potential interest. The most commonly used parameters include the average treatment effect (ATE), the ATE on the treated (ATT), and the ATE on the untreated (ATU), defined as

$$\begin{aligned} \tau_{\text{ATE}} &= \mathbb{E}(\tau_i) = \mathbb{E}\{Y_i(1) - Y_i(0)\} \\ \tau_{\text{ATT}} &= \mathbb{E}(\tau_i|T = 1) = \mathbb{E}\{Y_i(1) - Y_i(0)|T = 1\} \\ \tau_{\text{ATU}} &= \mathbb{E}(\tau_i|T = 0) = \mathbb{E}\{Y_i(1) - Y_i(0)|T = 0\} \end{aligned}$$

These parameters may also vary with a vector of covariates, X, in which case the parameters have an analogous representation conditional on a particular value of X.[1]

For nonrandom treatment assignment, selection into treatment may follow one of two general paths: 1) selection on observed variables, also referred to as unconfoundedness or the CIA (Rubin 1974; Heckman and Robb 1985); and 2) selection on unobserved variables. Under the CIA, selection into treatment is random conditional on covariates, X, and the average effect of the treatment can be obtained by comparing outcomes of individuals in the two treatment states with identical values of the covariates. This approach often uses propensity-score methods to reduce the dimensionality problem arising when X is a high-dimensional vector (Rosenbaum and Rubin 1983), with the propensity score denoted by $P(X_i) = \Pr(T_i = 1|X_i)$.

If the CIA fails to hold, then the estimated treatment effects relying on the CIA are biased. Following Heckman and Navarro-Lozano (2004) and Black and Smith (2004), we denote the potential outcomes as $Y(0) = g_0(X) + \varepsilon_0$ and $Y(1) = g_1(X) + \varepsilon_1$, where $g_0(X)$ and $g_1(X)$ are the deterministic portions of the outcome variable in the control and treatment groups, respectively, and where $(\varepsilon_0, \varepsilon_1)$ are the corresponding error terms. We also denote the latent treatment variable by $T^* = h(X) - u$, where $h(X)$ represents the deterministic portion of T^*, and u denotes the error term. The observed treatment, T, is therefore equal to 1 if $T^* > 0$ and 0 otherwise. Finally, we denote by δ the difference in the residuals of the potential outcomes, $\delta = \varepsilon_0 - \varepsilon_1$.

1. More formally, the coefficient measures the treatment effect, adjusting for a simultaneous linear change in the covariates, X, rather than being conditional on a specific value of X. We thank an anonymous referee for highlighting this point.

Assuming δ and u are jointly normally distributed, the bias can be derived as

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

where $\rho_{\delta u}$ is the correlation between δ and u, ρ_{0u} is the correlation between ε_0 and u, σ_0 is the standard deviation of ε_0, σ_δ is the standard deviation of δ, and ϕ and Φ are the normal probability density function and cumulative distribution function, respectively.

When the CIA fails, consistent estimation of the treatment effect of interest requires an alternative technique robust to selection on unobservables. This is difficult because obtaining a consistent point estimate of a measure of the treatment effect typically requires an exclusion restriction, which is unavailable in many situations. The proposed `bmte` command presents a series of treatment-effects estimators designed to estimate the average effects of treatment when appropriate exclusion restrictions are unavailable, exploiting the functional form of the bias in (1). Below we briefly present five of the estimators implemented by the `bmte` command.

2.1 The MB estimator

This technique relates generally to the normalized IPW estimator of Hirano and Imbens (2001), given by

$$\widehat{\tau}_{\text{IPW,ATE}} = \frac{\sum_{i=1}^{N} \dfrac{Y_i T_i}{\widehat{P}(X_i)}}{\sum_{i=1}^{N} \dfrac{T_i}{\widehat{P}(X_i)}} - \frac{\sum_{i=1}^{N} \dfrac{Y_i(1-T_i)}{1-\widehat{P}(X_i)}}{\sum_{i=1}^{N} \dfrac{(1-T_i)}{1-\widehat{P}(X_i)}} \tag{2}$$

where $\widehat{P}(X_i)$ is an estimate of the propensity score obtained using a probit model.

Under the CIA, the IPW estimator in (2) provides an unbiased estimate of τ_{ATE}. When this assumption fails, the bias for the ATE follows the closed functional form in (1), with similar expressions for the ATT and ATU. The MB estimator aims to minimize the bias by estimating (2) using only observations with a propensity score close to the bias-minimizing propensity score, denoted by P^*. Using P^* effectively limits the observations included in the estimation of the IPW treatment effects to minimize the inherent bias when the CIA fails. We denote by Ω the set of observations ultimately included in the estimation. In general, however, P^* and Ω are unknown. Therefore, the MB estimator estimates P^* and Ω to minimize the bias in (1) by using Heckman's BVN selection model, the details of which are provided in Millimet and Tchernis (2013).

The MB estimator of the ATE is formally given by

$$\widehat{\tau}_{\text{MB,ATE}}(P^*) = \frac{\sum_{i\in\Omega} \dfrac{Y_i T_i}{\widehat{P}(X_i)}}{\sum_{i\in\Omega} \dfrac{T_i}{\widehat{P}(X_i)}} - \frac{\sum_{i\in\Omega} \dfrac{Y_i(1-T_i)}{1-\widehat{P}(X_i)}}{\sum_{i\in\Omega} \dfrac{(1-T_i)}{1-\widehat{P}(X_i)}} \tag{3}$$

where $\Omega = \{i|P(X_i) \in C(P^*)\}$, and $C(P)$ denotes a neighborhood around P. Following Millimet and Tchernis (2013), the MB estimator defines $C(P^*)$ as $C(P^*) = \{\widehat{P}(X_i)|\widehat{P}(X_i) \in (\underline{P}, \overline{P})\}$, where $\underline{P} = \max(0.02, P^* - \alpha_\theta)$, $\overline{P} = \min(0.98, P^* + \alpha_\theta)$, and $\alpha_\theta > 0$ is the smallest value such that at least θ percent of both the treatment and control groups are contained in Ω. Specific values of θ are specified within the `bmte` command, with smaller values reducing the bias at the expense of higher variance. The MB estimator trims observations with propensity scores above and below specific values, regardless of the value of θ. These threshold values can be specified within the `bmte` command options. Obtaining Ω does not require the use of Heckman's BVN selection model when the focus is on the ATT or ATU, because P^* is known to be one-half in these cases (Black and Smith 2004).

If the user is sensitive to potential deviations from the normality assumptions underlying Heckman's BVN model, the MB estimator and other estimators can be extended appropriately (Millimet and Tchernis 2013). Such adjustments are included as part of the `bmte` command, denoted by the Edgeworth-expansion versions of the relevant estimators.

2.2 The BC approach

Estimation of the error correlation structure using Heckman's BVN model immediately introduces the possibility of a BC version of each estimator. Specifically, estimates of the bias of the MB estimator of the ATE, denoted by $\widehat{B_{\text{ATE}}(P^*)}$, can be derived from the two-stage BVN model. The estimated bias can then be applied as an adjustment to the standard IPW treatment-effects estimate.

The MB bias-corrected (MB-BC) estimator for the ATE is then given by

$$\widehat{\tau}_{\text{MB-BC,ATE}}(P^*) = \widehat{\tau}_{\text{MB,ATE}}(P^*) - \widehat{B_{\text{ATE}}(P^*)} \tag{4}$$

where the corresponding estimators for the ATT and ATU follow. With heterogeneous treatment effects, the MB-BC estimator changes the parameter being estimated. To identify the correct parameter of interest, the `bmte` command first estimates the MB-BC estimator in (4) conditional on the propensity score, $P(X)$, and then estimates the (unconditional) ATE by taking the expectation of this over the distribution of X in the population (or subpopulation of the treated). The resulting BC estimator is given by

$$\widehat{\tau}_{\text{BC,ATE}} = \widehat{\tau}_{\text{IPW,ATE}} - \sum_i \widehat{B_{\text{ATE}}\{\widehat{P}(X_i)\}} \tag{5}$$

where again the corresponding estimators for the ATT and ATU follow.

2.3 BVN selection

Briefly, Heckman's BVN selection model adopts a two-stage approach: 1) estimate the probability of treatment, $\Phi(X_i\widehat{\gamma})$, using a standard probit model with binary treatment as the dependent variable; and 2) estimate via OLS the following second-stage outcome equation,

$$\begin{aligned} Y_i = X_i\beta_0 + X_iT_i(\beta_1 - \beta_0) + \beta_{\lambda 0}(1 - T_i)\left\{\frac{\phi\left(X_i\widehat{\gamma}\right)}{1 - \Phi(X_i\widehat{\gamma})}\right\} \\ + \beta_{\lambda 1}T_i\left\{\frac{-\phi(X_i\widehat{\gamma})}{\Phi\left(X_i\widehat{\gamma}\right)}\right\} + \eta_i \end{aligned} \tag{6}$$

where $\phi(\cdot)/\Phi(\cdot)$ is the inverse Mills ratio, and η is an independent and identically distributed error term with constant variance and zero conditional mean. With this approach, the estimated ATE is given by

$$\widehat{\tau}_{\text{BVN,ATE}} = \overline{X}\left(\widehat{\beta}_1 - \widehat{\beta}_0\right) \tag{7}$$

Similar expressions are available for the ATT and ATU.[2]

2.4 CF approach

Heckman's BVN selection model is a special case of the CF approach. The idea is to devise a function where the treatment assignment is no longer correlated with the error term in the outcome equation once it is included, as outlined nicely in Heckman, LaLonde, and Smith (1999) and Navarro (2008). Specifically, consider the outcome equation

$$Y_i(t) = \alpha_t + g_t(X_i) + \mathbb{E}(\varepsilon_t|X_i, T_i = t) + \eta_{it}, \quad t = 0, 1$$

Approximating $\mathbb{E}(\varepsilon_t|X, T = t)$ with a polynomial in $P(X)$ yields

$$Y_i(t) = (\alpha_t + \pi_{t0}) + g_t(X_i) + \sum_{s=1}^{S}\pi_{ts}P(X_i)^s + \eta_{it}, \quad t = 0, 1$$

where S is the order of the polynomial. The following equation is then estimable via OLS:

$$\begin{aligned} Y_i = (\alpha_0 + \pi_{00})(1 - T_i) + (\alpha_1 + \pi_{10})T_i + X_i\beta_0 + X_iT_i(\beta_1 - \beta_0) \\ + \sum_{s=1}^{S}\pi_{0s}(1 - T_i)P(X_i)^s + \sum_{s=1}^{S}\pi_{1s}T_iP(X)^s + \eta_i \end{aligned} \tag{8}$$

2. Depending on one's dataset and specific application, it may not be meaningful to evaluate all covariates at their means. Therefore, when interpreting the treatment-effects estimates, the user should check that the data support the use of $\overline{X}$. We are grateful to an anonymous referee for clarifying this important point.

As is clear from (8), α_t and π_{t0} are not separately identified; however, because the selection problem disappears in the tails of the propensity score, it follows that the CF becomes zero and that the intercepts from the potential-outcome equations are identified using observations in the extreme end of the support of $P(X)$. After one estimates the intercept terms, the ATE and ATT are given by

$$\widehat{\tau}_{\text{CF,ATE}} = (\widehat{\alpha}_1 - \widehat{\alpha}_0) + \overline{X}\left(\widehat{\beta}_1 - \widehat{\beta}_0\right) \text{ and} \tag{9}$$

$$\widehat{\tau}_{\text{CF,ATT}} = (\widehat{\alpha}_1 - \widehat{\alpha}_0) + \overline{X}_1\left(\widehat{\beta}_1 - \widehat{\beta}_0\right) + \widehat{\mathbb{E}(\varepsilon_1 - \varepsilon_0|T_i = 1)} \tag{10}$$

where

$$\widehat{\mathbb{E}(\varepsilon_0|T_i = 1)} = -\left\{\sum_{s=1}^{S} \widehat{\pi}_{0s}\overline{P(X)_0^s}\right\}\left\{\frac{1-\overline{P(X)}}{\overline{P(X)}}\right\} \text{ and}$$

$$\widehat{\mathbb{E}(\varepsilon_1|T_i = 1)} = -\sum_{s=1}^{S} \widehat{\pi}_{1s} + \sum_{s=1}^{S} \widehat{\pi}_{1s}\overline{P(X)_1^s}$$

and where $\overline{P(X)}$ is the overall mean propensity score, and $\overline{P(X)_t}$, $t = 0, 1$, is the mean propensity score in group t.

2.5 Klein and Vella (2009) estimator

Unlike the CF approach, which relies on observations at the extremes of the support of $P(X)$, the Klein and Vella (2009) (KV) estimator attempts to identify the treatment effect by using more information from the middle of the support. Our implementation of the KV estimator relies on a similar functional form assumption to the BVN estimator in the absence of heteroskedasticity but effectively induces a valid exclusion restriction in the presence of heteroskedasticity. Specifically, denote the latent treatment by $T^* = X\gamma - u^*$, where $u^* = S(X)u$, $S(X)$ is an unknown positive function, and $u \sim N(0, 1)$. Here $S(X)$ is intended to allow for a general form of heteroskedasticity in the treatment effects.

In this case, the probability of receiving the treatment conditional on X is given by

$$\Pr(T = 1|X) = \Phi\left\{\frac{X}{S(X)}\gamma\right\} \tag{11}$$

Assuming $S(X) = \exp(X\kappa)$, the parameters of (11) are estimable by maximum likelihood, with the log-likelihood function given by[3]

$$\ln \mathcal{L} = \sum\nolimits_i \left[\ln \Phi\left\{\frac{X\gamma}{\exp(X\kappa)}\right\}\right]^{T_i}\left(\ln\left[1 - \Phi\left\{\frac{X\gamma}{\exp(X\kappa)}\right\}\right]\right)^{1-T_i} \tag{12}$$

3. Our functional form assumption, $S(X) = \exp(X\kappa)$, is a simplification made to compare the KV estimator and the other estimators available with the `bmte` command. For more details on the KV estimator and alternative functional forms for $S(X)$, see Klein and Vella (2009).

where the element of κ corresponding to the intercept is normalized to zero for identification. The maximum likelihood estimates are then used to obtain the predicted probability of treatment, $\widehat{P(X)}$, which may be used as an instrument for T in (6), excluding the selection correction terms.

3 The bmte command

3.1 Syntax

The `bmte` command implements the above MB, BC, BVN, CF, and KV estimators as well as the traditional OLS and IPW estimators. The syntax for the `bmte` command is

`bmte` *depvar indepvars* [*if*] [*in*], `group(`*varname*`)` [`ee hetero theta(#) psvars(`*indepvars*`) kv(`*indepvars*`) cf(#) pmin(#) pmax(#) psate(#) psatt(#) psatu(#) psateee(#) psattee(#) psatuee(#) saving(`*filename*`) replace bs reps(#) fixp`]

3.2 Specification

The `bmte` command requires the user to specify an outcome variable, *depvar*, at least one independent variable, and a treatment assignment variable, `group()`. Additional independent variables are optional. The command also uses Stata commands `hetprob` and `ivreg2` (Baum, Schaffer, and Stillman 2003, 2004, 2005). The remaining options of the `bmte` command are detailed below.

3.3 Options

`group(`*varname*`)` specifies the treatment assignment variable. `group()` is required.

`ee` indicates that the Edgeworth-expansion versions of the MB, BVN, and BC estimators be included in addition to the original versions of each respective estimator. The Edgeworth expansion is robust to deviations from normality in Heckman's BVN selection model.

`hetero` allows for heterogeneous treatment effects, with ATE, ATT, and ATU estimates presented at the mean level of each independent variable.

`theta(#)` denotes the minimum percentage such that both the treatment and control groups have propensity scores in the interval $(\underline{P}, \overline{P})$ from (3). Multiple values of `theta()` are allowed (for example, `theta(5 25)`, for 5% and 25%). Each value will form a different estimated treatment effect using the MB and MB-BC estimators.

`psvars(`*indepvars*`)` denotes the list of regressors used in the estimation of the propensity score. If unspecified, the list of regressors is assumed to be the same as the original covariate list.

`kv(`*indepvars*`)` denotes the list of independent variables used to model the variance in the `hetprob` command. Like the `psvars()` option, the list of `kv()` regressors is assumed to be the same as the original covariate list if not explicitly specified.

`cf(#)` specifies the order of the polynomial used in the CF estimator. The default is `cf(3)`.

`pmin(#)` and `pmax(#)` specify the minimum and maximum propensity scores, respectively, included in the MB estimator. Observations with propensity scores outside this range will be automatically excluded from the MB estimates. The defaults are `pmin(0.02)` and `pmax(0.98)`.

`psate(#)`–`psatuee(#)` specify the fixed propensity-score values (specific to each treatment effect of interest) to be used as the bias-minimizing propensity scores in lieu of estimating the values within the program itself.

`saving(`*filename*`)` indicates where to save the output.

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

`bs` and `reps(#)` specify that 95% confidence intervals be calculated by bootstrap using the percentile method and the number of replications in `reps(#)`. The default is `reps(100)`.

`fixp` is an option for the bootstrap command that, when specified, estimates the bias-minimizing propensity score $\{P^*(X)\}$ and applies this estimate across all bootstrap replications rather than reestimating at each replication.

4 Example

Following Millimet and Tchernis (2013), we provide an application of the `bmte` command to the study of the U.S. school breakfast program (SBP). Specifically, we seek causal estimates of the ATEs of SBP on child health. The data are from the Early Childhood Longitudinal Study—Kindergarten Class of 1998–1999 and are available for download from the *Journal of Applied Econometrics* Data Archive.[4] We provide estimates of the effect of SBP on growth rate in body mass index from first grade to the spring of third grade.

4. http://qed.econ.queensu.ca/jae/datasets/millimet001/.

We first define global variable lists `XVARS` and `HVARS` and limit our analysis to third grade students only. `XVARS` are the covariates used in the OLS estimation as well as in the calculation of the propensity score. `HVARS` are the covariates used in the KV estimator (that is, the variables that enter into the heteroskedasticity portion of the `hetprob` command).

```
. infile using millimettchernissbpdictionary.dct
  (output omitted)
. global XVARS gender age white black hispanic city suburb
> neast mwest south wicearly wicearlymiss momafb momafbmiss
> momft mompt momnw momeda momedb momedc momedd momede ses
> sesmiss bweight bweightmiss hfoodb hfoodbmiss books
> booksmiss momafb2 ses2 bweight2 books2 age2 z1-z22
. global HVARS ses age south city
```

We then estimate the effect of SBP participation in the first grade (`break1`) on body mass index growth (`clbmi`) by using the `bmte` command. In our application, we specify a θ of 5% and 25%, and we estimate bootstrap confidence intervals using 250 replications. We also specify the `ee` option, asking that the results include the Edgeworth-expansion versions of the relevant estimators. The resulting Stata output is as follows:

```
. bmte clbmi $XVARS if grade==3, g(break1) t(5 25) ee psv($XVARS) bs reps(250)
> kv($HVARS)
```

Theta	ATE	ATT	ATU
OLS	0.007 0.003, 0.011]	0.007 [0.003, 0.011]	0.007 [0.003, 0.011]
IPW	0.009 0.005, 0.014]	0.006 [0.002, 0.012]	0.011 [0.005, 0.012]
MB			
0.05	0.015 -0.008, 0.022]	-0.000 [-0.011, 0.014]	-0.000 [-0.011, 0.014]
0.25	0.005 -0.002, 0.011]	0.005 [-0.001, 0.011]	0.004 [-0.003, 0.009]
MB-EE			
0.05	0.014 0.005, 0.033]	0.009 [0.003, 0.023]	0.020 [0.002, 0.036]
0.25	0.013 0.005, 0.020]	0.005 [-0.001, 0.012]	0.013 [0.004, 0.022]
CF	0.048 -0.043, 0.120]	0.077 [-0.021, 0.159]	0.035 [-0.050, 0.107]
		F = 5.677 p = 0.000	

KV-IV	-0.008 -0.037, 0.022]	-0.008 [-0.037, 0.022] F = 133.462 p = 0.000 LR = 27.393 p = 0.000	-0.008 [-0.037, 0.022]
BVN	-0.017 -0.046, 0.012]	-0.003 [-0.021, 0.015]	-0.021 [-0.059, 0.015]
BVN-EE	0.230 0.052, 0.330]	0.134 [0.033, 0.187]	0.310 [0.070, 0.187]
MB-BC			
0.05	-0.007 -0.050, 0.018]	-0.019 [-0.055, 0.020]	-0.026 [-0.053, 0.002]
0.25	-0.017 -0.048, 0.011]	-0.014 [-0.049, 0.022]	-0.022 [-0.047, 0.002]
MB-BC-EE			
0.05	0.070 -0.039, 0.220]	0.212 [0.024, 0.304]	0.268 [0.009, 0.393]
0.25	0.069 -0.048, 0.215]	0.208 [0.020, 0.299]	0.261 [0.000, 0.388]
P*	0.672 0.167, 0.963]	0.500 [0.500, 0.500]	0.500 [0.500, 0.500]
P*-EE	0.033 0.020, 0.943]	0.787 [0.728, 0.956]	0.141 [0.020, 0.399]
BC-IPW	-0.018 -0.048, 0.012]	-0.014 [-0.048, 0.022]	-0.004 [-0.059, 0.022]
BC-IPW-EE	0.229 0.050, 0.331]	0.313 [0.070, 0.439]	1.063 [0.269, 0.439]

Here we focus on the general structure and theme of the output. For a thorough discussion and interpretation of the results, see Millimet and Tchernis (2013). As indicated by the section headings, the output presents results for the ATE, ATT, and ATU using basic OLS and IPW treatment-effects estimates as well as each of the MB (3), MB-BC (4), BC (5), BVN (7), CF [(9) and (10)], and KV [(11), (12), and (6)] estimators. Below each estimate is the respective 95% confidence interval.

As discussed in Millimet and Tchernis (2013), separate MB and MB-BC estimates are presented for each value of θ specified in the `bmte` command (in this case, 5% and 25%). The results for the CF estimator also include a joint test of significance of all covariates in the OLS step of the CF estimator (8). Similarly, the KV results include a test for weak instruments (the Cragg–Donald Wald F statistic and p-value) as well as a likelihood-ratio test for heteroskedasticity based on the results of `hetprob`. Also included in the `bmte` output is the estimated bias-minimizing propensity score.

We wish to reemphasize two points regarding the appropriate interpretation of results. First, the MB estimators will generally alter the interpretation of the parameter being estimated. Thus they may estimate a parameter considered to be uninteresting. Therefore, researchers should pay attention to the value of P^* as well as the attributes of observations with propensity scores close to this value. Second, none of the estimators considered here match the performance of a traditional IV estimator, although IV may also change the interpretation of the parameter being estimated.

5 Remarks

Despite advances in the program evaluation literature, treatment-effects estimators remain severely limited when the CIA fails and when valid exclusion restrictions are unavailable. Following the methodology presented in Millimet and Tchernis (2013), we propose and describe a new Stata command (`bmte`) that provides a range of treatment-effects estimates intended to estimate the average effects of the treatment when the CIA fails and appropriate exclusion restrictions are unavailable.

Importantly, the `bmte` command provides results that are useful across a range of alternative assumptions. For example, if the CIA holds, the IPW estimator provided by the `bmte` command yields an unbiased estimate of the causal effects of treatment. The MB estimator then offers a robustness check, given its comparable performance when the model is correctly specified or overspecified and its improved performance if the model is underspecified. If, however, the CIA does not hold, the `bmte` command provides results that are appropriate under strong functional form assumptions, either with homoskedastic (BVN or CF) or heteroskedastic (KV) errors, or under less restrictive functional form assumptions (BC). As illustrated in our example application to the U.S. SBP, the breadth of estimators implemented with the `bmte` command provides a broad picture of the average causal effects of the treatment across a variety of assumptions.

6 References

Baum, C. F., M. E. Schaffer, and S. Stillman. 2003. Instrumental variables and GMM: Estimation and testing. *Stata Journal* 3: 1–31.

———. 2004. Software updates: Instrumental variables and GMM: Estimation and testing. *Stata Journal* 4: 224.

———. 2005. Software updates: Instrumental variables and GMM: Estimation and testing. *Stata Journal* 5: 607.

Black, D. A., and J. Smith. 2004. How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of Econometrics* 121: 99–124.

Heckman, J., and S. Navarro-Lozano. 2004. Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics* 86: 30–57.

Heckman, J., and R. Robb, Jr. 1985. Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics* 30: 239–267.

Heckman, J. J. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5: 475–492.

———. 1979. Sample selection bias as a specification error. *Econometrica* 47: 153–161.

Heckman, J. J., R. J. LaLonde, and J. A. Smith. 1999. The economics and econometrics of active labor market programs. In *Handbook of Labor Economics*, ed. O. Ashenfelter and D. Card, vol. 3A, 1865–2097. Amsterdam: Elsevier.

Hirano, K., and G. W. Imbens. 2001. Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes Research Methodology* 2: 259–278.

Klein, R., and F. Vella. 2009. A semiparametric model for binary response and continuous outcomes under index heteroscedasticity. *Journal of Applied Econometrics* 24: 735–762.

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

Navarro, S. 2008. Control function. In *The New Palgrave Dictionary of Economics*, ed. S. N. Durlauf and L. E. Blume, 2nd ed. London: Palgrave Macmillan.

Rosenbaum, P. R., and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41–55.

Rubin, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66: 688–701.

About the authors

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

Rusty Tchernis is an associate professor of economics in the Andrew Young School of Policy Studies at Georgia State University. He is also a research associate at the National Bureau of Economic Research and a Research Fellow at the Institute for the Study of Labor. His primary areas of research are applied econometrics, health economics, and labor economics. Before becoming a faculty member at the Andrew Young School, he was an assistant professor in the Department of Economics at Indiana University and a postdoctoral research fellow in

the Department of Health Care Policy at Harvard Medical School. He received his PhD in economics from Brown University.

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