The bmte Command: Methods for the Estimation of Treatment Effects when Exclusion Restrictions are Unavailable

Ian McCarthy
Baylor Health Care System
Dallas, TX
ianmccarthy.econ@gmail.com

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

Rusty Tchernis
Georgia State University, Atlanta, GA
IZA, Bonn, Germany
NBER, Cambridge, MA
rtchernis@gsu.edu

Abstract. We present a new Stata command, **bmte** (bias-minimizing treatment effects), which implements two new estimators proposed in Millimet and Tchernis (2012) designed to estimate the effect of treatment when there exists selection on unobserved variables 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: 1) Heckman's two-step estimator (Heckman 1976, 1979); 2) a control function approach outlined in Heckman et al. (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 estimators alongside pre-existing 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 (2012).

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), there exist a variety of appropriate methodologies to estimate the causal effects of the treatment. If instead individuals self-select on the basis of unobserved variables (selection on unobserved variables), estimation of treatment effects is more difficult.

When confronted with selection on unobserved variables, the most common

1

empirical approach is to rely on an instrumental variable (IV); however, if credible instruments are unavailable, a small number of approaches now exist that attempt to estimate the effects of the treatment without an exclusion restriction. This paper introduces a new Stata command, **bmte**, which implements two recent estimators proposed in Millimet and Tchernis (2012) designed to estimate treatment effects when there exists selection on unobserved variables and appropriate exclusion restrictions are unavailable:

1. The *minimum-biased* (MB) estimator: This estimator implements a search for the observations among which the bias in the treatment effect estimate of interest is minimized. 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 (i.e., independence between treatment assignment and potential outcomes, conditional on observed variables), the MB estimator is unbiased. Otherwise, the MB estimator will tend to minimize the bias among estimators that rely on the CIA. Importantly, the MB estimator changes the parameter being estimated because of the restricted estimation sample.

2. The *bias-corrected* (BC) estimator: This estimator relies on the two-step estimator of Heckman's bivariate normal (BVN) selection model in order 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 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: 1) Heckman's BVN estimator; 2) a control function approach outlined in Heckman et al. (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 estimators alongside pre-existing 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 (2012). 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 include the average treatment effect (ATE), the ATE on the treated (ATT), and the ATE on the untreated (ATU), defined as:

$$\tau_{ATE} = \mathbb{E}[\tau_i] = \mathbb{E}[Y_i(1) - Y_i(0)] \tag{1}$$

$$\tau_{ATT} = \mathbb{E}[\tau_i|T = 1] = \mathbb{E}[Y_i(1) - Y_i(0)|T = 1] \tag{2}$$

$$\tau_{ATU} = \mathbb{E}[\tau_i|T = 0] = \mathbb{E}[Y_i(1) - Y_i(0)|T = 0]. \tag{3}$$

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]

In the case of non-random treatment assignment, selection into treatment may follow one of two general paths: 1) selection on observed variables, also referred to as the CIA or *unconfoundedness* (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 employs 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 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 were $(\varepsilon_0, \varepsilon_1)$ 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. 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, $\delta = \varepsilon_0 - \varepsilon_1$.

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

$$B_{ATE}[P(X)] = -\left\{\rho_{0u}\sigma_0 + [1 - P(X)]\rho_{\delta u}\sigma_\delta\right\} \frac{\phi(h(X))}{\Phi(h(X))[1 - \Phi(h(X))]}, \tag{4}$$

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 where ϕ and Φ are the normal PDF and CDF, respectively.

---

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

When the CIA fails, consistent estimation of the treatment effect of interest requires an alternative technique robust to selection on unobervables. The difficulty in this case is that obtaining a consistent *point estimate* of a measure of the treatment effect typically requires an exclusion restriction, which is not available in many situations. The proposed **bmte** command presents a series of treatment effects estimators designed to to estimate the average effects of treatment when appropriate exclusion restrictions are unavailable, exploiting the functional form of the bias in (4). Below, we present briefly five of the estimators implemented by the **bmte** command.

2.1 The Minimum-Biased Estimator

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

$$\hat{\tau}_{IPW,ATE} = \frac{\sum_{i=1}^{N} \frac{Y_i T_i}{\hat{P}(X_i)}}{\sum_{i=1}^{N} \frac{T_i}{\hat{P}(X_i)}} - \frac{\sum_{i=1}^{N} \frac{Y_i (1-T_i)}{1-\hat{P}(X_i)}}{\sum_{i=1}^{N} \frac{(1-T_i)}{1-\hat{P}(X_i)}}, \tag{5}$$

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

Under the CIA, the IPW estimator in (5) provides an unbiased estimate of τ_{ATE}. When this assumption fails, the bias for the ATE follows the closed functional form in (4), with similar expressions for the ATT and ATU. The MB estimator aims to minimize the bias by estimating (5) using only observations with a propensity score in a neighborhood around the bias-minimizing propensity score, denoted by P^*. The use of P^* effectively limits the observations included in the estimation of the IPW treatment effects in such a way as 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 not known. The MB estimator therefore estimates P^* and Ω so as to minimize the bias in equation 4 using Heckman's BVN selection model, the details of which are provided in Millimet and Tchernis (2012).

The MB estimator of the ATE is formally given by

$$\hat{\tau}_{MB,ATE}[P^*] = \frac{\sum_{i \in \Omega} \frac{Y_i T_i}{\hat{P}(X_i)}}{\sum_{i \in \Omega} \frac{T_i}{\hat{P}(X_i)}} - \frac{\sum_{i \in \Omega} \frac{Y_i (1-T_i)}{1-\hat{P}(X_i)}}{\sum_{i \in \Omega} \frac{(1-T_i)}{1-\hat{P}(X_i)}}, \tag{6}$$

where $\Omega = \{i | \hat{P}(X_i) \in C(P^*)\}$ and $C(P)$ denotes a neighborhood around P. Following Millimet and Tchernis (2012), the MB estimator defines $C(P^*)$ as $C(P^*) = \{\hat{P}(X_i) | \hat{P}(X_i) \in (\underline{P}, \overline{P})\}$, 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. Note that the MB estimator trims observations with propensity scores above and below specific values, regardless of the

4

value of θ. These threshold values can be specified within the **bmte** command options. Note also that obtaining Ω does not require the use of Heckman's BVN selection model when the focus is on the ATT or ATU, since P^* is known to be one-half in these cases (Black and Smith 2004).

If the user is particularly 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 2012). Such adjustments are included as part of the **bmte** command, denoted by the Edgeworth expansion (EE) versions of the relevant estimators.

2.2 The Bias-Corrected Approach

Estimation of the error correlation structure using Heckman's BVN model immediately gives rise to the possibility of a bias-corrected version of each estimator. Specifically, estimates of the *bias* of the MB estimator of the ATE, denoted by $\widehat{B_{ATE}}[P^*]$, 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 minimum bias bias-corrected estimator, denoted by MB-BC, for the ATE is then given by

$$\hat{\tau}_{MB-BC,ATE}[P^*] = \widehat{\tau}_{MB,ATE}[P^*] - \widehat{B_{ATE}[P^*]}, \tag{7}$$

where the corresponding estimators for the ATT and ATU follow. With heterogeneous treatment effects, the MB-BC estimator changes the parameter being estimated. In order to identify the correct parameter of interest, the **bmte** command first estimates the MB-BC estimator in (7) conditional on the propensity score, $P(X)$, and then estimates the (unconditional) average treatment effect by taking the expectation of this over the distribution of X in the population (or sub-population of treated). The resulting bias-corrected (BC) estimator is given by

$$\hat{\tau}_{BC,ATE} = \widehat{\tau}_{IPW,ATE} - \sum_i \widehat{B_{ATE}[\hat{P}(X_i)]}, \tag{8}$$

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, $\Phi(X_i\hat{\gamma})$, using a standard probit model with binary treatment as the dependent variable; and 2) estimate via OLS the following second-stage outcome equation:

$$
\begin{aligned}
Y_i &= X_i\beta_0 + X_iT_i(\beta_1 - \beta_0) + \beta_{\lambda 0}(1 - T_i)\left[\frac{\phi(X_i\hat{\gamma})}{1 - \Phi(X_i\hat{\gamma})}\right] \\
&\quad + \beta_{\lambda 1}T_i\left[\frac{-\phi(X_i\hat{\gamma})}{\Phi(X_i\hat{\gamma})}\right] + \eta_i
\end{aligned}
\tag{9}
$$

5

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

$$\hat{\tau}_{BVN,ATE} = \overline{X}\left(\widehat{\beta}_1 - \widehat{\beta}_0\right). \tag{10}$$

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

2.4 Control Function Approach

Heckman's BVN selection model is a special case of the control function (CF) approach. The idea is to devise a function such that treatment assignment is no longer correlated with the error term in the outcome equation upon its inclusion, as outlined nicely in Heckman et al. (1999) and Navarro (2008). Specifically, consider the outcome equation

$$Y_i(t) = \alpha_t + g_t(X_i) + \mathbb{E}[\varepsilon_t | X_i, T_i = t] + \eta_{it}, t = 0, 1. \tag{11}$$

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

$$Y_i(t) = (\alpha_t + \pi_{t0}) + g_t(X_i) + \sum_{s=1}^{S} \pi_{ts} P(X_i)^s + \eta_{it}, \ t = 0, 1, \tag{12}$$

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

$$\begin{aligned} Y_i &= (\alpha_0 + \pi_{00})(1 - T_i) + (\alpha_1 + \pi_{10})T_i + X_i\beta_0 + X_iT_i(\beta_1 - \beta_0) \quad (13) \\ &+ \sum_{s=1}^{S} \pi_{0s}(1 - T_i)P(X_i)^s + \sum_{s=1}^{S} \pi_{1s}T_iP(X)^s + \eta_i. \end{aligned}$$

As is clear from (13), α_t and π_{t0} are not separately identified; however, since the selection problem disappears in the tails of the propensity score, it follows that the control function becomes zero and the intercepts from the potential outcome equations are identified based on observations in the extreme end of the support of $P(X)$. After estimating the intercept terms, the ATE and ATT are given by

$$\hat{\tau}_{CF,ATE} = (\widehat{\alpha}_1 - \widehat{\alpha}_0) + \overline{X}\left(\widehat{\beta}_1 - \widehat{\beta}_0\right) \text{ and} \tag{14}$$

$$\hat{\tau}_{CF,ATT} = (\widehat{\alpha}_1 - \widehat{\alpha}_0) + \overline{X}_1\left(\widehat{\beta}_1 - \widehat{\beta}_0\right) + \mathbb{E}[\varepsilon_1 - \widehat{\varepsilon_0 | T_i} = 1], \tag{15}$$

---

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

where

$$
\begin{aligned}
\widehat{\mathbb{E}[\varepsilon_0|T_i=1]} &= -\left[\sum_{s=1}^{S}\widehat{\pi}_{0s}\overline{P(X)_0^s}\right]\left(\frac{1-\overline{P(X)}}{\overline{P(X)}}\right) \text{ and} \\
\widehat{\mathbb{E}[\varepsilon_1|T_i=1]} &= -\sum_{s=1}^{S}\widehat{\pi}_{1s}+\sum_{s=1}^{S}\widehat{\pi}_{1s}\overline{P(X)_1^s},
\end{aligned}
$$

and where $\overline{P(X)_t}$ 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 & Vella (2009) Estimator

Unlike the CF approach, which relies on observations at the extremes of the support of $P(X)$, the KV estimator attempts to identify the treatment effect utilizing 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 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

$$
\Pr(T=1|X)=\Phi\left(\frac{X}{S(X)}\gamma\right). \tag{16}
$$

Assuming $S(X)=\exp(X\kappa)$, the parameters of (16) are estimable by maximum likelihood (ML), with the log-likelihood function given by:[3]

$$
\ln\mathcal{L}=\sum_i\left[\ln\Phi\left(\frac{X\gamma}{\exp(X\kappa)}\right)\right]^{T_i}\left\{\ln\left[1-\Phi\left(\frac{X\gamma}{\exp(X\kappa)}\right)\right]\right\}^{1-T_i}, \tag{17}
$$

where the element of κ corresponding to the intercept is normalized to zero for identification. The ML estimates are then used to obtain the predicted probability of treatment, $\widehat{P(X)}$, which may be used as an instrument for T in equation (9) 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

---

[3]Our functional form assumption, $S(X)=\exp(X\kappa)$, is a simplification made for comparison purposes between 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).

`bmte` *depvar indepvars* $\lceil$*if*$\rceil$ $\lceil$*in*$\rceil$, `group(`*varname*`)` $\lceil$<u>`theta`</u>`(#)`
   <u>`psv`</u>`ars(`*indepvars*`) kv(`*indepvars*`) ee` <u>`hetero`</u> `cf(#) pmin(#)`
   `pmax(#) bs reps(#) fixp psate(#) psatt(#) psatu(#)`
   `psateee(#) psattee(#) psatuee(#)` <u>`saving`</u>`(`*string*`) replace`$\rceil$

3.2 Specification and Options

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 utilizes Stata commands `hetprob` and `ivreg2` (Baum et al. 2003, 2004, 2005). The remaining options of the **bmte** command are detailed below:

`theta(#)` denotes the minimum percentage such that both the treatment and control groups have propensity scores in the interval $(\underline{P}, \overline{P})$ from equation (6). The user can specify multiple values for θ (e.g., `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 left 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 in Stata. Like the `psvars` option, the list of `kv` regressors is assumed to be the same as the original covariate list if not explicitly specified.

`ee` specifies that Edgeworth-expansion versions of the relevant estimators also be provided.

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

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

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

`bs` specifies that 95% confidence intervals be calculated by bootstrap using the percentile method based on the number of replications in `reps(#)`. If `reps` is

unspecified, the program defaults to 100 replications.

`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 re-estimating at each replication.

`psate(#)` - `psatuee(#)` allow the user to specify 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.

4   Example

Following Millimet and Tchernis (2012), 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 average treatment effects of SBP on child health. The data are from the Early Childhood Longitudinal Study - Kindergarten Class of 1998-99 (ECLS-K) 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 BMI from first grade to spring of third grade.

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 for the calculation of the propensity score. `HVARS` are the covariates used in the KV estimator (i.e., the variables that enter into the heteroskedasticity portion of the `hetprob` command).

```
. 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 BMI growth (*clbmi*) using the **bmte** command. In our application, we specify a θ of 5% and 25%, and we estimate bootstrap confidence intervals based on 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 follows:

---

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

```
. 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	0.007	0.007
	[0.003, 0.011]	[0.003, 0.011]	[0.003, 0.011]
IPW	0.009	0.006	0.011
	[0.005, 0.014]	[0.001, 0.012]	[0.006, 0.012]
MB			
0.05	0.015	-0.000	-0.000
	[-0.010, 0.020]	[-0.011, 0.011]	[-0.012, 0.011]
0.25	0.005	0.005	0.004
	[-0.002, 0.011]	[-0.001, 0.011]	[-0.003, 0.010]
MB-EE			
0.05	0.014	0.009	0.020
	[0.008, 0.032]	[0.001, 0.023]	[0.003, 0.035]
0.25	0.013	0.005	0.013
	[0.004, 0.020]	[-0.002, 0.012]	[0.005, 0.021]
CF	0.048	0.077	0.035
	[-0.037, 0.115]	[-0.015, 0.152]	[-0.055, 0.105]
		F = 5.677	
		p = 0.000	
KV-IV	-0.008	-0.008	-0.008
	[-0.040, 0.027]	[-0.040, 0.027]	[-0.040, 0.027]
		F = 133.461	
		p = 0.000	
		LR = 27.393	
		p = 0.000	
BVN	-0.017	-0.003	-0.021
	[-0.047, 0.012]	[-0.020, 0.014]	[-0.060, 0.014]
BVN-EE	0.230	0.134	0.310
	[0.058, 0.331]	[0.043, 0.187]	[0.085, 0.187]
MB-BC			
0.05	-0.007	-0.019	-0.026
	[-0.049, 0.014]	[-0.055, 0.015]	[-0.052, -0.000]
0.25	-0.017	-0.014	-0.022
	[-0.050, 0.011]	[-0.048, 0.021]	[-0.049, 0.003]
MB-BC-EE			
0.05	0.070	0.212	0.268
	[-0.033, 0.230]	[0.047, 0.301]	[0.028, 0.429]
0.25	0.069	0.208	0.261
	[-0.037, 0.220]	[0.042, 0.291]	[0.021, 0.419]
P*	0.672	0.500	0.500
	[0.105, 0.967]	[0.500, 0.500]	[0.500, 0.500]
P*-EE	0.033	0.787	0.141
	[0.020, 0.944]	[0.737, 0.930]	[0.020, 0.427]

BC-IPW	-0.018	-0.014	-0.004
	[-0.049, 0.012]	[-0.050, 0.020]	[-0.061, 0.020]

BC-IPW-EE	0.229	0.313	1.063
	[0.058, 0.330]	[0.100, 0.431]	[0.395, 0.431]

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 (2012). 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 (equation 6), MB-BC (equation 7), BC (equation 8), BVN (equation 10), CF (equations 14 and 15), and KV (equations 16, 17, and 9) estimators. Below each estimate is the respective 95% confidence interval.

As discussed in the text, 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 (equation 13). 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.

It is also worth re-emphasizing two points regarding the appropriate interpretation of results. First, the MB estimators will generally alter the interpretation of the parameter being estimated. As a result, they may estimate a parameter considered to be uninteresting. Researchers should therefore 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 (2012), 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 useful across a range of alternative assumptions. For example, if the CIA holds, the IPW estimator provided by the **bmte** command provides 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 over-specified, but improved performance if the model is under-specified. If, however, the CIA does not hold, the **bmte** command provides results appropriate under strong

11

functional form assumptions, either with homoskedastic (BVN or CF) or heteroskedastic (KV) errors, or less restrictive functional form assumptions (BC). As illustrated in our example application to U.S. school breakfast program, 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., M. 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. A. Smith. 2004. How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of Econometrics* 121(1): 99–124.

Heckman, J. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models 5(4): 475–492.

———. 1979. Sample selection bias as a specification error. *Econometrica* 47(1): 153–161.

Heckman, J., R. J. LaLonde, and J. A. Smith. 1999. The economics and econometrics of active labor market programs. *Handbook of labor economics* 3: 1865–2097.

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(1): 30–57.

Heckman, J., and R. Robb. 1985. Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics* 30(1): 239–267.

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(3): 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(5): 735–762.

Millimet, D. L., and R. Tchernis. 2012. Estimation of Treatment Effects Without an Exclusion Restriction: with an Application to the Analysis of the School Breakfast Program. *Journal of Applied Econometrics* . URL `http://dx.doi.org/10.1002/jae.2286`.

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

Rosenbaum, P. R., and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1): 41–55.

Rubin, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66(5): 688–701.