The Stata Journal

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

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

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

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 (2014)
14, Number 1, pp. 191–217

Estimating marginal treatment effects using parametric and semiparametric methods

Scott Brave
Federal Reserve Bank of Chicago
Chicago, IL
sbrave@frbchi.org

Thomas Walstrum
University of Illinois at Chicago
and
Federal Reserve Bank of Chicago
Chicago, IL
twalstrum@frbchi.org

Abstract. We describe the new command `margte`, which computes marginal and average treatment effects for a model with a binary treatment and a continuous outcome given selection on unobservables and returns. Marginal treatment effects differ from average treatment effects in instances where the impact of treatment varies within a population in correlation with unobserved characteristics. Both parametric and semiparametric estimation methods can be used with `margte`, and we provide evidence from a Monte Carlo simulation for when each is preferable.

Keywords: st0331, margte, locpoly2, etregress, movestay, marginal treatment effect, average treatment effect, generalized Roy model, local instrumental variables

1 Introduction

The estimation of marginal treatment effects (MTEs) is an approach used in empirical research when the impact of a treatment is thought to vary within a population in correlation with unobserved characteristics. For instance, Carneiro, Heckman, and Vytlacil (2011) use this framework to measure the differential returns to education for individuals whose unobserved characteristics make it more likely for them to pursue higher education. Other applications include Doyle (2007) and Brinch, Mogstad, and Wiswall (2012), who use this framework to estimate the effects of foster care and family size, respectively, on the long-term outcomes of children.

To illustrate the difference between MTEs and average treatment effects (ATEs), consider the classic return-to-education model in labor economics. These models usually begin with the Mincer equation describing the log-level of wages for a collection of individuals,

$$\log(\text{wage}_i) = \alpha + \beta \text{enroll}_i + \gamma_1 \text{exp}_i + \gamma_2 \text{exp}_i^2 + u_i \tag{1}$$

where enroll_i is a binary variable for whether an individual enrolled in a postsecondary school, exp_i is the subsequent work experience of the individual, and u_i represents unobservable wage determinants. If u_i is an independent, identically distributed (i.i.d.) normal random variable, such that $\text{Cov}(u_i, \text{enroll}_i) = 0$, then ordinary least squares provides an unbiased estimate of β, which represents the average return to postsecondary education conditional on experience.

 st0331

Suppose, however, that the true model of wage determination is

$$\log(\text{wage}_i) = \alpha + \beta\text{enroll}_i + \gamma_1\text{exp}_i + \gamma_2\text{exp}_i^2 + \delta\text{mot}_i + \nu_i \tag{2}$$

where mot_i captures the motivation of an individual. We would expect that, on average, people who are more motivated earn more and are more likely to have attended a postsecondary school. Unfortunately, a person's motivation is unobservable to the econometrician. This means that we cannot know whether people who attended a postsecondary school have higher earnings because of their education or because they are more motivated. In this case, the u_i in (1) is equal to $\delta\text{mot}_i + \nu_i$, and the ordinary least-squares estimate of β is biased.

The above example is an instance of what is often referred to as selection on unobservables. A standard solution to this problem is to find an instrumental variable. In our example, such a variable must be correlated with postsecondary school enrollment but uncorrelated with u_i. Using a valid instrument with two-stage least squares provides an unbiased estimate of β if the true empirical model is (2) but the econometrician only observes the variables in (1).

Suppose, instead, that the model of wage determination is

$$\log(\text{wage}_i) = \alpha + \beta\text{enroll}_i + \gamma_1\text{exp}_i + \gamma_2\text{exp}_i^2 + \delta\text{mot}_i + \theta\text{enroll}_i \times \text{mot}_i + u_i \tag{3}$$

Now the average return to education varies throughout the population according to $\beta + \theta \times \text{mot}_i$. This instance of the Mincer equation is said to exhibit selection on returns (see Carneiro, Heckman, and Vytlacil [2011]). Another way to present it is to begin with (1) but to treat β as a random variable such that

$$\beta_{(1)} = \beta_{(3)} + \theta_{(3)}\text{mot}_i + \epsilon$$

where ϵ is an i.i.d. unobserved random variable and the subscripts in parentheses reference the equation numbers in the text. If the scale of motivation is normalized such that $E(\text{mot}_i) = 0$, then $\beta_{(1)}$ is the average return to postsecondary education, or ATE.

If $\theta_{(3)} = 0$, then there is no selection on returns and two-stage least-squares estimation of (1) is unbiased even though $\beta_{(1)}$ is random. However, in our example, we would expect θ to be positive such that more motivated people get more out of their time spent in school. It is this dependence, called essential heterogeneity, that makes it relevant to examine the marginal return to postsecondary education, or MTE, for individuals with varying levels of motivation.

The presence of selection on returns necessitates that we explicitly model the treatment decision. This is because, as is made clear in our example, any variable that is correlated with the decision to enroll in a postsecondary school is also correlated with the unobserved interaction between that decision and motivation. Heckman, Urzua, and Vytlacil (2006) show that the selection probability into treatment, or the propensity score, is a valid instrument given selection on unobservables and selection on returns, and it can be used to identify both ATEs and MTEs.

Returning to our example, the propensity score p_i is the expected value of enroll_i conditional on observable variables z_i that help to explain enrollment.

$$p_i = E(\text{enroll}_i | z_i) \tag{4}$$

Instrumental variables for selection on unobservables are included in z_i. Taking the expectation of log wages, equation (3), conditional on exp_i and z_i then implies

$$E\{\log(\text{wage}_i)|\text{exp}_i, z_i\} = K(p_i, \text{mot}_i) = K(p_i)$$

where $K(p_i, \text{mot}_i)$ is a function of the propensity score and motivation that is conditional on exp_i and z_i. The inclusion of a valid instrument in z_i ensures that p_i and mot_i are uncorrelated, so we can simply write this function as $K(p_i)$. The MTE is then defined as the derivative of $K(p_i)$. The MTE can be evaluated for each p in a range of values defined by (4).

$$\text{MTE} = \frac{\partial E\{\log(\text{wage}_i)|\text{exp}_i, z_i\}}{\partial p} = K'(p)$$

By integrating MTE over p, we arrive again at the ATE.

$$\text{ATE} = \int_p \text{MTE}\ dp$$

Thus in our example, the MTE tell us how much an individual's wage increases when there is a small increase in the propensity score or, equivalently, how much higher the wages of an individual that is on the margin of treatment can be expected to be by inducing them to enroll in college via the instruments in z_i. The presence of an instrumental variable in the treatment decision model for p_i ensures that the reason for this increase is unrelated to motivation. Also, because the propensity score equals the unobserved propensity to not enroll in a postsecondary school for indifferent individuals, we can capture the marginal return to education for varying levels of motivation.

Notice that if motivation were observable in our example, the MTE is the same as $\beta_{(1)}$.

$$\begin{aligned} E\{\log(\text{wage}_i)|\text{exp}_i, z_i\} &= \alpha + \beta p_i + \gamma_1 \text{exp}_i + \gamma_2 \text{exp}_i^2 + \delta \text{mot}_i + \theta p_i \times \text{mot}_i \\ \text{MTE} &= \frac{\partial E\{\log(\text{wage}_i)|\text{exp}_i, z_i\}}{\partial p_i} = \beta + \theta \text{mot}_i \end{aligned}$$

If more motivated people are more likely to have enrolled in a postsecondary school and their return to education is higher, the distribution of MTE over values of the propensity score will show this. However, if there is more than one unobservable factor at play in the decision to enroll in a postsecondary school (perhaps in addition to motivation, persistence is also important), it is impossible to distinguish between the effects of the two. The most we can say is that if motivated and persistent people are more likely to enroll and their return to education is higher, they will exhibit larger MTEs.

In the next section, we describe how to use the propensity score to identify MTE and ATE. Then we show how this model can be estimated using a new command, `margte`, which nests the existing command `etregress` (see [TE] **etregress**) and the `movestay` command of Lokshin and Sajaia (2004) in its options. Unlike the other commands, `margte` produces estimates of both the MTE and ATE by using either parametric or semiparametric methods for estimating $K(p_i)$. Finally, we evaluate the appropriate uses of our estimators with a Monte Carlo simulation designed to test their identification assumptions.

2 Marginal treatment effects

In this section, we motivate the derivation and estimation of MTE within the statistical framework provided by the generalized Roy model.

2.1 The generalized Roy model

As noted in Heckman (2010), the generalized Roy model is an example of a broader class of treatment-effects models which jointly model a continuous outcome and its binary treatment. MTE is a parameter of the generalized Roy model. Our description of the model below closely follows that found in Heckman, Urzua, and Vytlacil (2006).

The potential outcomes (Y_0, Y_1) of a treatment $D = (0,1)$ are assumed to depend linearly upon observable variables $\mathbf{X}$ and unobservables (U_0, U_1). The decision process for the treatment indicator is posed as a function of observables $\mathbf{Z}$ and unobservables V, and linked to the observed outcome Y_D through the latent variable I.

$$\begin{aligned}
Y_D &= (1-D)Y_0 + DY_1 \\
Y_1 &= \alpha_1 + \mathbf{X}\boldsymbol{\beta}_1 + U_1 \\
Y_0 &= \alpha_0 + \mathbf{X}\boldsymbol{\beta}_0 + U_0 \\
I &= \mathbf{Z}\boldsymbol{\gamma} - V \qquad (5) \\
D &= \begin{cases} 1 & \text{if } I > 0 \\ 0 & \text{if } I \le 0 \end{cases} \qquad (6)
\end{aligned}$$

The model is identified either through parametric restrictions on U_0, U_1, and V or by including variables in $\mathbf{Z}$ that satisfy the following constraints: $\text{Cov}(\mathbf{Z}, U_0) = \mathbf{0}$, $\text{Cov}(\mathbf{Z}, U_1) = \mathbf{0}$, and $\boldsymbol{\gamma} \neq \mathbf{0}$.

Written in this way, the generalized Roy model encompasses both of the treatment-effects models fit by the commands `etregress` and `movestay`. For instance, if $\boldsymbol{\Sigma}$ is the variance–covariance matrix of unobservables and

$$(U_0, U_1, V) \sim N(\mathbf{0}, \boldsymbol{\Sigma})$$

we obtain an identical representation to the endogenous switching regression model described in Lokshin and Sajaia (2004) and fit by the command `movestay`. Furthermore, by also restricting that

$$\alpha_0 = \alpha_1$$
$$\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$$
$$\sigma_1^2 = \sigma_0^2$$

where σ_i^2 represents the variance of U_i in $\boldsymbol{\Sigma}$, we obtain the treatment-effects model described in Maddala (1983) and fit by the Stata command `etregress`.

The first assumption restricts the functional form of the variance–covariance matrix of the unobservable determinants of Y and D by using what is known as a "control function" approach to identification. The second set of more restrictive assumptions ensures that the expectation of Y conditional on $\mathbf{X}$ as well as the marginal effect of $\mathbf{X}$ on Y is independent of treatment status. Both commands make it possible to parametrically estimate ATE as $E(Y_1 - Y_0)$ by using maximum likelihood methods, or also by using a two-step consistent estimator with `etregress`.

Unlike `movestay` and `etregress`, the `margte` command produces estimates of both MTE and ATE by using either parametric or semiparametric methods. To see how this is possible, consider the following. Without loss of generality, we can redefine (5) as

$$I > 0 \Leftrightarrow \mathbf{Z}\gamma > V \Leftrightarrow F_V(\mathbf{Z}\gamma) > F_V(V) \Leftrightarrow P(\mathbf{Z}) > U_D$$

where F_V is the cumulative distribution function of V, often called a link function, and D is the treatment status of an individual. Written in this way, $P(\mathbf{Z})$, the propensity score, denotes the selection probability of treatment, while U_D is a uniformly distributed random variable between 0 and 1 representing the propensity not to be treated.

The MTE is the marginal benefit of treatment ($D = 1$) conditional on $\mathbf{X}$ and the propensity not to be treated (U_D), as shown in Bjorklund and Moffitt (1987):

$$\text{MTE} \equiv E(Y_1 - Y_0 | \mathbf{X} = \mathbf{x}, U_D = u_D) \tag{7}$$

This contrasts with the ATE, which captures the average benefit associated with treatment conditional on $\mathbf{X}$:

$$\text{ATE} \equiv E(Y_1 - Y_0 | \mathbf{X} = \mathbf{x})$$

Heckman and Vytlacil (2001b) and Heckman, Urzua, and Vytlacil (2006) show that the ATE can be constructed as a weighted average of the MTE by integrating over U_D.

The estimated propensity score $\widehat{P}(z) = \Pr(\mathbf{Z}\boldsymbol{\gamma} > V | \mathbf{Z} = \mathbf{z})$ allows us to define the range of U_D over which MTE is identified. Given $\widehat{P}(\mathbf{z})$, the following conditional expectations of Y by observed treatment status form the basis of the parametric estimation procedure supported by `margte`.

$$E\{Y | \mathbf{X} = \mathbf{x}, P(\mathbf{Z}) = p, D = 1\} = \alpha_1 + \mathbf{x}\boldsymbol{\beta}_1 + E\{U_1 | \mathbf{X} = \mathbf{x}, P(\mathbf{Z}) = p, D = 1\} \tag{8}$$
$$E\{Y | \mathbf{X} = \mathbf{x}, P(\mathbf{Z}) = p, D = 0\} = \alpha_0 + \mathbf{x}\boldsymbol{\beta}_0 + E\{U_0 | \mathbf{X} = \mathbf{x}, P(\mathbf{Z}) = p, D = 0\} \tag{9}$$

Following Heckman and Vytlacil (2001a), (8)–(9) can be rewritten as

$$E\{Y|\mathbf{X}=\mathbf{x},P(\mathbf{Z})=p\}=\alpha_0+\mathbf{x}\boldsymbol{\beta}_0+(\alpha_1-\alpha_0)p+\{\mathbf{x}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\}p+K(p) \tag{10}$$

$$K(p)=E\{U_0|P(\mathbf{Z})=p\}+E\{U_1-U_0|P(\mathbf{Z})=p\}p$$

to arrive at a semiparametric representation of the conditional expectation of Y also capable of being estimated by `margte`. Semiparametric estimators of the MTE, however, require an additional identification assumption on the support of the estimated propensity score, which is discussed further below.

2.2 Parametric estimators of the MTE

By assuming that $(U_0, U_1, V) \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the variance–covariance matrix of the three unobservables, we can estimate the MTE over the range of $P(\mathbf{Z})$, that is, $(0,1)$. The propensity score is generated from a probit model where $P(\mathbf{Z}) = \Phi\{(\mathbf{Z}\boldsymbol{\gamma})/(\sigma_V)\}$ and Φ is the cumulative normal distribution. Following standard practice with the probit model, we normalize its scale such that $\sigma_V = 1$.

Using the definition for the MTE in (7) where $U_D = \Phi(V)$,

$$\text{MTE}(\mathbf{X}=\mathbf{x},U_D=u_D)=(\alpha_1-\alpha_0)+\mathbf{x}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)+(\rho_1-\rho_0)\Phi^{-1}(u_D)$$

such that ρ_i, $i=(0,1)$, corresponds to the element of $\boldsymbol{\Sigma}$ containing the covariance between U_i and V. Estimation of the parameters of the MTE then follows from the linear regressions implied by (8)–(9) using

$$\begin{aligned} E\{U_1|\mathbf{X}=\mathbf{x},P(\mathbf{Z})=p,D=1\} &= -\rho_1\frac{\phi(p)}{\Phi(p)p} \\ E\{U_0|\mathbf{X}=\mathbf{x},P(\mathbf{Z})=p,D=0\} &= \rho_0\frac{\phi(p)}{\Phi(p)(1-p)} \end{aligned}$$

where the two fractions in the above expressions are the inverse Mills ratios.

It is possible to partially relax the assumption of joint normality, which also allows $P(\mathbf{Z})$ to be fit by another probability model. In this case, the command `margte` allows the propensity score to be fit as a linear probability or logit model.[1]

Given an estimate of $P(\mathbf{Z})$, (10) can be written as

$$E\{Y|\mathbf{X}=\mathbf{x},P(\mathbf{Z})=p\}=\alpha_0+\mathbf{x}\boldsymbol{\beta}_0+(\alpha_1-\alpha_0)p+\mathbf{x}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)p+\sum_{i=1}^{\vartheta}\phi_i p^i \tag{11}$$

where $K(p)$ is approximated by a polynomial in p of chosen degree ϑ.

1. The linear probability model should be used with caution given that its range for $P(\mathbf{Z})$ is not constrained to be $(0,1)$.

Here the MTE is defined as the partial derivative of the conditional expectation of Y with respect to $P(\mathbf{Z})$,

$$\frac{\partial E\{Y|\mathbf{X}=\mathbf{x},P(\mathbf{Z})=p\}}{\partial p} = (\alpha_1-\alpha_0)+\mathbf{x}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)+\frac{\partial K(p)}{\partial p} \tag{12}$$

such that

$$\text{MTE}\{\mathbf{X}=\mathbf{x},P(\mathbf{Z})=p\} = (\alpha_1-\alpha_0)+\mathbf{x}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)+\sum_{i=1}^{\vartheta} i\phi_i p^{i-1}$$

Its parameters are estimated by the linear regression implied by (11).[2]

2.3 Semiparametric estimators of the MTE

Heckman, Urzua, and Vytlacil (2006b) describe two semiparametric estimation strategies for the MTE. Identification in both of these instances depends crucially on the common support assumption for the propensity score, which requires that there exist positive frequencies of $\widehat{P}(z)$ in the range of $(0,1)$ for individuals that do $(D=1)$ and do not $(D=0)$ receive treatment. Verifying that a common support exists requires first specifying a probability model, or link function F_V, for the propensity score. Given an estimate of the propensity score, the range of common support is determined by `margte` before estimation of the MTE, and a histogram is presented to capture the result.

Drawing on (12), the semiparametric estimators of the MTE are computed according to

$$\text{MTE}\{\mathbf{X}=\mathbf{x},P(\mathbf{Z})=p\} = \frac{\partial E\{Y|\mathbf{X}=\mathbf{x},P(\mathbf{Z})=p\}}{\partial p} = \mathbf{x}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)+\frac{\partial K(p)}{\partial p} \tag{13}$$

where, without any further assumptions on $K(p)$, the estimation of the last term requires the use of nonparametric techniques for local derivatives.[3]

One approach to estimating (13), known as local instrumental variables (LIV), is to first run local linear regressions of $\mathbf{X}$, $\mathbf{X}\times P(\mathbf{Z})$, and Y on $P(\mathbf{Z})$ at every observed value of $\widehat{P}(\mathbf{Z})$ to obtain estimated residuals $\widehat{e}_Y$, $\widehat{\mathbf{e}}_\mathbf{X}$, and $\widehat{\mathbf{e}}_{\mathbf{X}\times P}$. By then regressing $\widehat{e}_Y$ on $\widehat{\mathbf{e}}_\mathbf{X}$ and $\widehat{\mathbf{e}}_{\mathbf{X}\times P}$, we arrive at an estimate of $\{\boldsymbol{\beta}_0,(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\}$ in a similar fashion to Heckman et al. (1998). Alternatively, similar to the way in which the assumption of joint normality can be relaxed in the parametric case, we can instead run the linear regression implied by (11) to obtain $\{\boldsymbol{\beta}_0,(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\}$.

2. The coefficient on $P(\mathbf{Z})$, ϕ_1, in this regression includes $\alpha_1-\alpha_0$ so that all the parameters of the MTE are identified.
3. The constant terms have been subsumed in the $\mathbf{X}$ matrix here and in what follows.

The remaining parameters of the MTE are then obtained from a local polynomial regression of

$$\widetilde{Y} = Y - \mathbf{X}\widehat{\beta}_0 - \{\mathbf{X}(\widehat{\beta_1 - \beta_0})\}P(\mathbf{Z})$$

on the common support of $P(\mathbf{Z})$ to arrive at an estimate of $\{\partial K(p)\}/(\partial p)$.[4]

Our semiparametric estimators use the Stata command `lpoly` (see [R] **lpoly**) to perform the local linear and polynomial regressions in the algorithm above as well as a modified version of its predecessor, `locpoly`, described in Gutierrez, Linhart, and Pitblado (2003). We modify the latter to store higher-order approximations as local derivatives, similar to Marsh (2006).[5]

3 The margte command

`margte`'s syntax preserves many of the stylistic features of `etregress`. The dependent and independent variables of the outcome equation are first listed, leaving out the binary treatment indicator variable. The treatment equation is then defined, listing the binary treatment variable (defined as 0s and 1s) and its covariates, in that order, with a separate option.

3.1 Syntax

Parametric normal model

`margte` *depvar*$_o$ *varlist*$_o$ [*if*] [*in*], `treatment(`*depvar*$_t$ *varlist*$_t$`)` [`first`
`link(`*string*`)` `common` `nocommongraph` `csbarwidth(`#`)` `xvalues(`#, #, ...`)`
`constraints(`#, #, ...`)` `mlikelihood` `mlopts(`*string*`)` `degree(`#`)`
`kernel(`*kernel*`)` `ybwidth(`#`)` `xbwidth(`#`)` `savepropensity` `noplot`
`plotci(`*string*`)` `noboot` `level(`#`)` `bca` `bsopts(`*string*`)`]

Parametric polynomial model

`margte` *depvar*$_o$ *varlist*$_o$ [*if*] [*in*], `treatment(`*depvar*$_t$ *varlist*$_t$`)` `polynomial(`#`)`
[`first` `link(`*string*`)` `common` `nocommongraph` `csbarwidth(`#`)`
`xvalues(`#, #, ...`)` `constraints(`#, #, ...`)` `mlikelihood` `mlopts(`*string*`)`
`degree(`#`)` `kernel(`*kernel*`)` `ybwidth(`#`)` `xbwidth(`#`)` `savepropensity` `noplot`
`plotci(`*string*`)` `noboot` `level(`#`)` `bca` `bsopts(`*string*`)`]

4. Local polynomial estimation techniques are explained in Fan and Gijbels (1996).
5. In keeping with the naming conventions already established, we call this command `locpoly2`. Only the ado-version of this command is currently supported. Additional details can be found in the accompanying help file for `locpoly2`.

Semiparametric LIV model

margte *depvar*$_o$ *varlist*$_o$ [*if*] [*in*], treatment(*depvar*$_t$ *varlist*$_t$) semiparametric [first link(*string*) common nocommongraph csbarwidth(#) xvalues(#, #, ...) constraints(#, #, ...) mlikelihood mlopts(*string*) degree(#) kernel(*kernel*) ybwidth(#) xbwidth(#) savepropensity noplot plotci(*string*) noboot level(#) bca bsopts(*string*)]

Semiparametric polynomial model

margte *depvar*$_o$ *varlist*$_o$ [*if*] [*in*], treatment(*depvar*$_t$ *varlist*$_t$) polynomial(#) semiparametric [first link(*string*) common nocommongraph csbarwidth(#) xvalues(#, #, ...) constraints(#, #, ...) mlikelihood mlopts(*string*) degree(#) kernel(*string*) ybwidth(#) xbwidth(#) savepropensity noplot plotci(*string*) noboot level(#) bca bsopts(*string*)]

3.2 Options

treatment(*depvar*$_t$ *varlist*$_t$) specifies the treatment equation that estimates the propensity score. The first variable in the list is the dependent variable and all following variables are the independent variables. The independent variable list should, in most cases, contain at least one variable that is not in the outcome equation. treatment() is required.

polynomial(#) specifies the degree of the polynomial in the propensity score used to fit $K(p)$ for the parametric and semiparametric polynomial models. If the option is not specified, margte will fit the parametric normal or semiparametric LIV model depending on whether the semiparametric option is also present. polynomial() is required when specifying a parametric polynomial model or a semiparametric polynomial model.

semiparametric specifies that the semiparametric LIV model or, when combined with the polynomial() option, the semiparametric polynomial model be fit. The option semiparametric is required when specifying a semiparametric LIV model or a semiparametric polynomial model.

first specifies that margte display the first-step estimates of the treatment equation before estimation. If the model is estimated by maximum likelihood, margte will display the output from movestay.

link(*string*) specifies the link function used in estimating the propensity score. It can be estimated using probit, logit, or the linear probability model (lpm). The default, link(probit), is also the only link function allowed if margte is fitting the parametric normal model.

`common` specifies that the common support be calculated and graphed. For U_D from 0.01 to 0.99 in increments of 0.01, a given value of U_D is in the common support if both treated and untreated observations are in the neighborhood $|U_D(\text{obs}) - U_D| < 0.005$. MTE is identified for semiparametric models only at values of U_D that have common support, thus `margte` automatically invokes this option if a semiparametric model is specified. Parametric models do not depend on common support for identification and by default do not invoke `common`.

`nocommongraph` suppresses the graph generated when the option `common` is specified.

`csbarwidth(#)` specifies the width of the bars in the common support graph. The default, `csbarwidth(0.1)`, gives a consistent appearance regardless of the graph's dimensions.

`xvalues(#, #, ...)` specifies the values of $varlist_o$ at which to calculate the MTE. The values must be separated by commas and follow the order of $varlist_o$. The default is to evaluate the MTE at the means of $varlist_o$.

`constraints(#, #, ...)` specifies linear constraints on the model's parameters. Type `help constraint` within Stata for more information.

`mlikelihood` fits the parametric normal model with maximum likelihood. When the option `mlikelihood` is specified, `margte` calls `movestay` and reformats the output to conform with the standard described here. To see the original output from `movestay`, specify option `first` as well. Postestimation hypothesis testing is allowed, but use caution because `e(V)` contains 0s when covariances are undefined. In such circumstances, `test` (see [R] **test**) may return an invalid answer.

`mlopts(`*string*`)` controls the maximization process in `movestay`. (Type `help movestay` and `help maximize` within Stata for details.) These options are seldom used.

`degree(#)` specifies the degree of the polynomial in the nonparametric regression of $\widetilde{Y}$ on $K(p)$ for the semiparametric LIV model. The regression provides $dK(p)/dp$, which is then used to calculate the MTE. The minimum degree allowed is 1. The default is `degree(2)`. The semiparametric polynomial model matches the degree to that specified in the `polynomial()` option. (Type `help locpoly2` within Stata for details.)

`kernel(`*kernel*`)` specifies the kernel function used in the nonparametric regressions of the semiparametric models. The default is `kernel(epanechnikov)`. (Type `help lpoly` within Stata for details.) `kernel(epan2)` is not allowed.

`ybwidth(#)` specifies the half-width of the kernel for $depvar_o$, that is, the width of the smoothing window around each point. The specified value applies to all nonparametric regressions involving $depvar_o$. If left unspecified, `margte` uses `lpoly`'s rule-of-thumb (ROT) bandwidth estimator.

`xbwidth(#)` specifies the half-width of the kernel for $varlist_o$, that is, the width of the smoothing window around each point. The specified value applies to all nonparametric regressions involving $varlist_o$. If left unspecified, `margte` uses `lpoly`'s ROT bandwidth estimator.

`savepropensity` saves the propensity score as the variable `p`. If any of the variables in memory are named `p`, then `margte` will return an error.

`noplot` suppresses the plot of the MTE.

`plotci(`*string*`)` specifies which confidence intervals to plot for the MTE from those provided by `bootstrap` (see [R] **bootstrap**). *string* can be `normal`, `percentile`, `bc`, and `bca`. Type `help bootstrap` within Stata for a detailed exposition on the differences between the options.

`noboot` turns off standard error bootstrapping. No closed-form solution for the standard error of the MTE exists. Because bootstrapping is computationally intensive, it may take a long time for `margte` to run.

`level(#)` specifies a confidence level for all standard errors. The default is `level(95)`.

`bca` computes acceleration for the bias-corrected confidence intervals. `bootstrap` automatically computes normal, percentile, and bias-corrected confidence intervals, but `bca` must be called separately because it is computationally intensive.

`bsopts(`*string*`)` specifies other `bootstrap` options. Useful options include `reps(#)` and `cluster(`*varlist*`)`. (Type `help bootstrap` within Stata for more information.)

4 Examples

In this section, we present example output of the `margte` command using simulated data from the generalized Roy model as described in section 2. To better illustrate the use of `margte` and to give interpretation to its output, we generate the simulated data with the accompanying command `margte_dgps` based on a model of the returns to education like the example in section 1.

4.1 Returns to education example

In keeping with our earlier example and the generalized Roy model in section 2.1, we take the treatment to be the decision of whether to enroll in a postsecondary school and the outcome to be the individual's future wages. More formally, the treatment equation consists of a binary decision model for postsecondary school enrollment, `enroll`, which depends on the sign of the continuous latent variable I where

$$
\begin{aligned}
I &= \gamma_0 + \gamma_1 \texttt{distCol} + \gamma_2 \texttt{momsEdu} - V \\
\texttt{enroll} &= \begin{cases} 1 & \text{if } I > 0 \\ 0 & \text{if } I \leq 0 \end{cases}
\end{aligned}
$$

and the variables distCol and momsEdu are drawn from the literature on the returns to education describing environmental factors impacting the enrollment decision. The variable distCol captures whether an individual grew up near a college, and momsEdu is the highest education level attained by the individual's mother. The outcome equation is a linear model for the log level of hourly wages, log(wage), such that

$$\begin{aligned}
\log(\text{wage}) &= (1-\text{enroll})\log(\text{wage})_0 + \text{enroll}\times\log(\text{wage})_1 \\
\log(\text{wage})_1 &= \alpha_1 + \beta_{11}\text{exp} + \beta_{12}\text{exp}^2 + \beta_{13}\texttt{momsEdu} + u_1 \\
\log(\text{wage})_0 &= \alpha_0 + \beta_{01}\text{exp} + \beta_{02}\text{exp}^2 + \beta_{03}\texttt{momsEdu} + u_0
\end{aligned}$$

where exp is work experience and u_0, u_1, and V are i.i.d. unobservable random variables. Subscripts of 1 reference those individuals who enrolled in a postsecondary school, while 0 subscripts reference those who did not.

This model's simulated data-generating process is specified in margte_dgps and is intended to capture the relationship between wages, college enrollment, work experience, mother's education, and distance to the nearest college, as described in chapter 5 of Wooldridge (2010). Experience, mother's education, and distance to the nearest college are generated as uniformly distributed random variables with means of 20 years, 12 years, and 25 miles, respectively. Their coefficients are then set such that the mean hourly wage rate for those who enroll in college is about \$34 per hour and is roughly \$25 per hour for those who do not.[6]

Notice that the binary decision model is equivalent to

$$\text{enroll} = \begin{cases} 1 & \text{if } \gamma_0 + \gamma_1\texttt{distCol} + \gamma_2\texttt{momsEdu} > V \\ 0 & \text{if } \gamma_0 + \gamma_1\texttt{distCol} + \gamma_2\texttt{momsEdu} \le V \end{cases}$$

where transforming the above by using the cumulative distribution of V, F_V, yields the propensity score function, $P(\texttt{distCol}, \texttt{momsEdu})$,

$$\begin{aligned}
F_V(\gamma_0 + \gamma_1\texttt{distCol} + \gamma_2\texttt{momsEdu}) &> F_V(V) \\
P(\texttt{momsEdu}, \texttt{distCol}) &> U_{\text{enroll}}
\end{aligned}$$

where U_{enroll} is a uniformly distributed random variable between 0 and 1 and serves as a standardized measure of a person's unobservable propensity not to enroll in a postsecondary school. Individuals who have a U_{enroll} close to 1 exhibit a large unobservable propensity to avoid postsecondary education. This will be an important feature to keep in mind when interpreting the output of the margte command.

Taking the expectation of the linear model for the log level of wages, we obtain the expected log wage conditional on experience and mother's education, where the conditional expectation for enroll is $P(\texttt{momsEdu}, \texttt{distCol}) = p$, the propensity score.

$$E\{\log(\text{wage})\} = (1-p)\times E\{\log(\text{wage})_0\} + p\times E\{\log(\text{wage})_1\}$$

6. Depending on the magnitude of the simulated unobservables, our calibration does not rule out negative wage rates. These instances occur at a rate of 1 in 6,000 observations, and they have no discernible effect on the results of our simulations.

Average and marginal returns to education conditional on experience and mother's education are then given by

$$
\begin{aligned}
\text{ATE} &= (\alpha_1 - \alpha_0) + (\beta_{11} - \beta_{01})\overline{\text{exp}} + (\beta_{12} - \beta_{02})\overline{\text{exp}^2} + (\beta_{13} - \beta_{03})\overline{\texttt{momsEdu}} \\
\text{MTE} &= \text{ATE} + E(u_1 - u_0)
\end{aligned}
$$

according to their definitions in section 2.1. The `margte` command allows the user to specify the values of experience, experience squared, and mother's education at which to calculate the conditional expectations above. The default is the mean of the variables and is used here and in all the examples that follow. The average return to education in our simulated data is 32 log points, or about 38%.

To estimate MTE, the researcher must make an assumption about the joint distribution of u_0, u_1, and V. The `margte` command allows V to have a marginal distribution that is normal (`probit`), logistic (`logit`), or uniform between 0 and 1 (`regress`). The marginal distributions of u_0 and u_1 are then determined by the choice of one of the four MTE estimators described in section 2. Each estimator calculates $E(u_1 - u_0)$ in a slightly different way, as described in section 2, based on the implied joint distribution of the model's unobservables.

The most common distributional assumption for the model's unobservables, and the one chosen by the Stata command `etregress` and the `movestay` command of Lokshin and Sajaia (2004), is the multivariate normal assumption. For our simulated data, we assume that u_0, u_1, and V are generated from a trivariate normal distribution with a known variance–covariance matrix, $\mathbf{\Sigma}$. Our calibration of $\mathbf{\Sigma}$ is then guided by the example in the introduction where individual motivation is an unobserved variable.

Consider the case where ν_0, ν_1, and ϵ_v are i.i.d. unobservable random variables

$$
\begin{aligned}
V &= \gamma_v \text{mot} + \epsilon_v \\
u_1 &= \gamma_1 \text{mot} + \nu_1 \\
u_0 &= \gamma_0 \text{mot} + \nu_0
\end{aligned}
$$

such that if $\gamma_1 > \gamma_0$, then more-motivated individuals have a higher return to education. If $\gamma_v \neq 0$, then the model exhibits selection on unobservables. Furthermore, if $\gamma_1 \neq 0$ or $\gamma_0 \neq 0$, so that motivation plays a role in the return to education, the model also exhibits selection on returns. We calibrate $\mathbf{\Sigma}$ such that the model above exhibits selection on unobservables and selection on returns.

The MTE then tells us how much higher or lower an individual's wage is expected to be given a small increase in the propensity score. The presence of an instrumental variable in the treatment equation means that the reason individuals with the same mother's education were induced to attend college (in this instance, happening to live closer to a college) is unrelated to unobserved motivation. Thus people at the margin of the treatment decision identify the MTE for everyone whose $U_{\text{enroll}} = p$. If more-motivated people are more likely to attend a postsecondary school and their return to education is higher, MTE is decreasing in U_{enroll}.

4.2 Parametric normal

We first estimate the parametric normal version of the MTE from section 2.2 with the `margte` command. Standard errors are calculated using Stata's `bootstrap` command. Because a closed-form solution for the standard error of the MTE does not exist, this is the preferred method of measuring the uncertainty surrounding its estimate. However, for the model's remaining parameters, bootstrapping may be overridden by specifying the option `mlikelihood`, which uses the maximum likelihood routine of `movestay` to estimate the parameters and their standard errors.

Without specifying the `mlikelihood` option, the generalized Roy model is fit in stages: first running the probit regression of `enroll` on `distCol` and `momsEdu` to obtain the propensity score p and the inverse Mills ratio k, followed by linear regressions of log wages on a constant, `exp`, `exp2`, `momsEdu`, and `k` for both treated and untreated cases. This two-step procedure is similar to that used by the Stata command `etregress`, but it is less restrictive in the sense that marginal effects and error covariances are not constrained to be equal for treated and untreated cases.

The output is organized by equation, displaying the parameter estimates for both the treated and untreated cases. Without specifying `first` as an option, the parameter estimates for the treatment equation are not displayed. Instead, its output is summarized by displaying the link function at the top of the table. For the purpose of exposition, we include this option here and report the estimates of the first-stage probit regression. In the next section, we examine the consequences for our MTE estimates of misspecifying the underlying model for the propensity score.

The example output below exhibits selection on unobservables, that is, the coefficients on the inverse Mills ratios (`rho1` and `rho0`) are statistically significant from 0 and statistically different from each other (`rho1` $-$ `rho0` < 0). The direction of selection is such that individuals who enrolled in a postsecondary school have unobservable characteristics that are negatively correlated with their unobservable wage determinants V, whereas those individuals who did not enroll exhibit a positive correlation. Our example also exhibits selection on returns, although this will not be readily apparent until we examine the estimated MTEs.

Before doing so, however, we describe the second-stage estimation results. Log hourly wages (`lwage`) are estimated to increase in work experience (`exp`) and mother's education (`momsEdu`) and decrease in experience squared (`exp2`) for both treated and untreated cases, with marginal effects that are slightly larger for the treated cases. Calculated at the mean of the independent variables, this implies an estimated average return to education (`E(Y1-Y0)@X`) of about 40%, in line with the true ATE of 38% that we used to simulate our data.

```
. margte lwage exp exp2 momsEdu, treatment(enroll momsEdu distCol) first
Iteration 0:   log likelihood = -3465.7259
Iteration 1:   log likelihood = -2497.5662
Iteration 2:   log likelihood =  -2492.236
Iteration 3:   log likelihood =  -2492.232
Iteration 4:   log likelihood =  -2492.232
Probit regression                               Number of obs   =       5000
                                                LR chi2(2)      =    1946.99
                                                Prob > chi2     =     0.0000
Log likelihood =  -2492.232                     Pseudo R2       =     0.2809

------------------------------------------------------------------------------
      enroll |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     momsEdu |   .1418175     .00918    15.45   0.000      .123825      .15981
     distCol |  -.0617553   .0016309   -37.87   0.000    -.0649517   -.0585589
       _cons |  -.1734692   .1112346    -1.56   0.119    -.3914849    .0445466
------------------------------------------------------------------------------
(running parametric_normal on estimation sample)
Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
Parametric Normal MTE Model                    Number of obs      =       5000
Treatment Model: Probit                        Replications       =         50

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Treated      |
         exp |   .1427322   .0046789    30.51   0.000     .1335618    .1519026
        exp2 |  -.0012403   .0001169   -10.61   0.000    -.0014693   -.0010112
     momsEdu |   .0727007   .0052843    13.76   0.000     .0623436    .0830578
           k |  -.1300469   .0294245    -4.42   0.000     -.187718   -.0723759
       _cons |   .3788084   .0869809     4.36   0.000     .2083291    .5492877
-------------+----------------------------------------------------------------
Untreated    |
         exp |   .1158638   .0037582    30.83   0.000     .1084979    .1232297
        exp2 |  -.0013701   .0000912   -15.02   0.000    -.0015488   -.0011913
     momsEdu |   .0483552   .0038956    12.41   0.000       .04072    .0559904
           k |   .2279336   .0215992    10.55   0.000        .1856    .2702672
       _cons |    .944638   .0598158    15.79   0.000     .8274013    1.061875
-------------+----------------------------------------------------------------
Mills        |
   rho1-rho0 |  -.3579805   .0356468   -10.04   0.000     -.427847   -.2881141
-------------+----------------------------------------------------------------
ATE          |
  E(Y1-Y0)@X |   .3378193   .0304774    11.08   0.000     .2780848    .3975538
------------------------------------------------------------------------------
```

MTE, because it is conditional on a given realization of the propensity score, is omitted from the table.[7] It is instead plotted separately (see figure 1) over the range of U_{enroll} (or U_D in the notation of section 2) consistent with the MTE estimator that is chosen [in this case $(0, 1)$]. Shaded error bands corresponding with a 95% confidence interval are also plotted, as is a dashed line for the ATE as a reference point.

7. Following Heckman, Urzua, and Vytlacil (2006b), `margte` calculates MTE over the range of values from 0.01 to 0.99 in increments of 0.01.

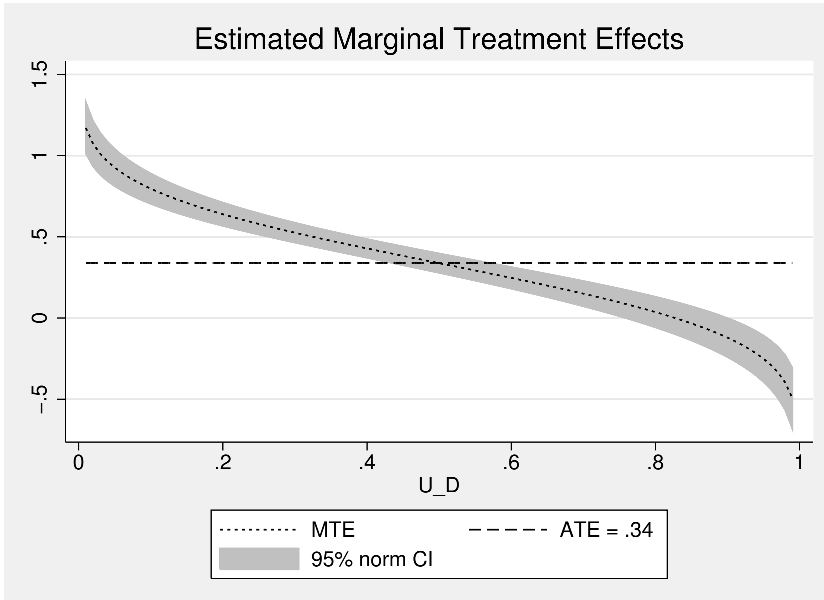

Figure 1. MTE over the common support of $P(\mathbf{Z})$

The estimated MTE in our example is decreasing in U_{enroll}, reflecting that the marginal return to education is increasing in the propensity of an individual to enroll in a postsecondary school. Our estimates range from a return of slightly more than 100% for individuals with the highest propensities for enrollment to roughly −50% for those with propensity scores near 0. The magnitude of the MTE results is consistent with our calibration of $\mathbf{\Sigma}$ (see appendix).

4.3 Parametric polynomial

The parametric polynomial version of the MTE from section 2.2 is obtained by specifying the desired degree of the propensity score used in estimating the expectation of log wages conditional on the propensity score. A link function other than probit may be specified. The example output below assumes a logit link function (estimation results not shown) and a fourth-order polynomial expansion of the propensity score. These options are not mutually exclusive: specifying an alternative to probit necessitates using a polynomial expansion of the propensity score.

The generalized Roy model is again fit in two stages: a first-stage logit regression to obtain the propensity score p, followed by a linear regression of log wages (`lwage`) on a constant, `exp`, `exp2`, `momsEdu`, and their interactions with p (`expXp`, `exp2Xp`, and `momsEduXp`) along with its polynomial terms (`p1`, `p2`, `p3`, and `p4`). Estimated coefficients and their standard errors are reported. Both the ATE and MTE are presented as above.

Log wages in this model are estimated to increase in work experience and mother's education, with their marginal effects also increasing in the propensity to have enrolled in a postsecondary school (that is, the coefficients on `expXp` and `momsEduXp` are both greater than 0). The fitted model also exhibits selection on unobservables and returns, because the linear and higher-order polynomial expansion terms of the propensity score are jointly statistically significant.[8]

This estimator performs almost as well as the parametric normal estimator at estimating the average and marginal returns to education in our example (see appendix). In the next section, we examine whether the reverse is true when the error structure of the model is nonnormal.

```
. margte lwage exp exp2 momsEdu, treatment(enroll momsEdu distCol) polynomial(4)
> link(logit) noplot
(running parametric_polynomial on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Parametric Polynomial MTE Model                Number of obs      =       5000
Treatment Model: Logit                         Replications       =         50

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Parameters   |
         exp |   .1187656   .0068602    17.31   0.000     .1053199    .1322112
        exp2 |  -.0014112   .0001628    -8.67   0.000    -.0017304   -.001092
     momsEdu |    .044385   .0084363     5.26   0.000     .0278501    .0609199
       expXp |   .0204441   .0098372     2.08   0.038     .0011635    .0397247
      exp2Xp |   .0002205    .000226     0.98   0.329    -.0002224    .0006634
   momsEduXp |   .0277618   .0159079     1.75   0.081     -.003417    .0589407
          p1 |   1.819088   .8074025     2.25   0.024     .2366086    3.401568
          p2 |  -7.799666   3.404652    -2.29   0.022    -14.47266   -1.12667
          p3 |   10.69073   5.089589     2.10   0.036      .715315    20.66614
          p4 |  -5.190432   2.568135    -2.02   0.043    -10.22388   -.1569803
       _cons |   .8440723   .1026043     8.23   0.000     .6429716    1.045173
-------------+----------------------------------------------------------------
ATE          |
  E(Y1-Y0)@X |   .3825122    .071144     5.38   0.000     .2430725    .5219518
------------------------------------------------------------------------------
```

8. Tests such as this one for essential heterogeneity are discussed in Heckman, Schmierer, and Urzua (2010).

4.4 Semiparametric LIV and polynomial

A semiparametric version of the generalized Roy model proceeds by specifying the option `semiparametric`. This fits the model with the LIV procedure described in section 2.3. If left unspecified, `margte` uses the default kernel and ROT bandwidth of `lpoly`. We allow for separate bandwidths to be specified for the local linear regressions used in estimation of the MTE with the options `xbwidth()` and `ybwidth()`. Specifying the option `ybwidth()` also determines the bandwidth used in the local polynomial regression estimated by `locpoly2`.[9] Only one kernel can be specified.

```
. margte lwage exp exp2 momsEdu, treatment(enroll momsEdu distCol)
> semiparametric xbwidth(.01) ybwidth(.2) noplot nocommongraph
(running semiparametric on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Semiparametric LIV MTE Model                    Number of obs     =      5000
Treatment Model: Probit                         Replications      =        50
```

lwage	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
Parameters						
exp	.1190972	.0054138	22.00	0.000	.1084864	.129708
exp2	-.0014177	.0001278	-11.10	0.000	-.0016681	-.0011673
momsEdu	.0438542	.0095762	4.58	0.000	.0250852	.0626231
expXp	.0200417	.0094316	2.12	0.034	.0015562	.0385272
exp2Xp	.000232	.0002201	1.05	0.292	-.0001993	.0006633
momsEduXp	.0278976	.0159799	1.75	0.081	-.0034224	.0592175
ATE						
E(Y1-Y0)@X	.3756926	.0629642	5.97	0.000	.252285	.4991002

The table output for this model resembles that for the parametric polynomial model. Regression coefficients and normal-based bootstrap standard errors are presented for the independent variables and their interactions with the propensity score. Advanced options of `margte` are available that instead allow for percentile as well as bias-corrected and accelerated standard errors. The ATE estimate is displayed at the bottom of the table, while the MTE estimates are again plotted separately in an accompanying figure (not shown).

When estimated using semiparametric methods, MTE is calculated only at values that fall within the common support of the first-stage estimates of the propensity score, shown graphically in figure 2. We recommend using this chart as a guide to gauging the reliability of MTE estimates over the range of values for U_D. Although the MTE is plotted as long as positive frequencies of treated and untreated cases exist, these frequencies often will be small and the MTE results should be appropriately discounted.

9. We highly recommend testing the sensitivity of the MTE estimates produced by `margte` to different choices for both bandwidths.

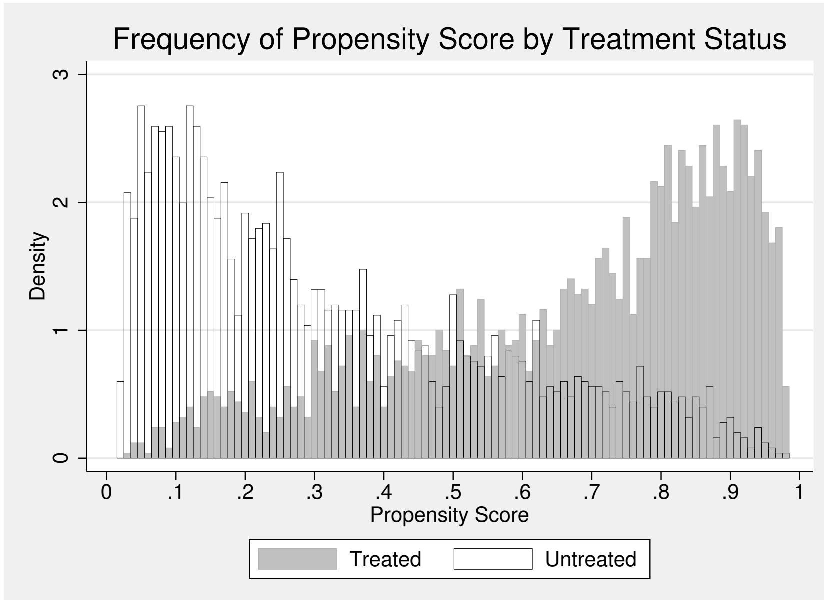

Figure 2. The common support of $P(\mathbf{Z})$

```
. margte lwage exp exp2 momsEdu, treatment(enroll momsEdu distCol)
> semiparametric polynomial(4) ybwidth(.2) noplot
(running semiparametric on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Semiparametric Polynomial MTE Model            Number of obs      =       5000
Treatment Model: Probit                        Replications       =         50
```

lwage	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
Parameters						
exp	.1186367	.0055981	21.19	0.000	.1076645	.1296088
exp2	-.0014083	.0001326	-10.62	0.000	-.0016683	-.0011484
momsEdu	.0447926	.0098547	4.55	0.000	.0254776	.0641075
expXp	.0207143	.0090892	2.28	0.023	.0028998	.0385287
exp2Xp	.0002149	.0002164	0.99	0.321	-.0002092	.000639
momsEduXp	.0273449	.016126	1.70	0.090	-.0042615	.0589512
p1	1.46855	.7444064	1.97	0.049	.0095401	2.92756
p2	-6.555712	3.016014	-2.17	0.030	-12.46699	-.6444331
p3	9.000845	4.373326	2.06	0.040	.429283	17.57241
p4	-4.405606	2.166042	-2.03	0.042	-8.650971	-.1602414
_cons	.8700232	.0989947	8.79	0.000	.6759971	1.064049
ATE						
E(Y1-Y0)@X	.3665734	.0951283	3.85	0.000	.1801254	.5530214

By further specifying a degree for the polynomial expansion of the propensity score, we arrive at the final estimator of the MTE. Our semiparametric estimates for this example are in line with their true values (see appendix), exhibiting both selection on unobservables and selection on returns. In the next section, we demonstrate the impact that a limited common support can have on the reliability of semiparametric estimates of MTE.

5 Monte Carlo simulation

In this section, we evaluate our four MTE estimators with Monte Carlo simulation techniques. We draw 5,000 different random samples, each with 5,000 observations, from the generalized Roy model according to the example of section 4. Using the options of `margte_dgps`, we specify both the model's assumed error structure (normal versus nonnormal) and which of the four MTE estimators to apply to each random sample. To approximate a nonnormal error structure, we simulate data for the conditional expectation of log wages by using a polynomial of degree 4 in the propensity score.

The tables in the appendix report the actual values of the generalized Roy model's parameters along with the sample mean and standard deviations of their estimates. Bias can be assessed in the tables by comparing the sample mean and actual values, while the standard deviations are equivalent to the root mean squared error between the actual and estimated values. All four estimators produce unbiased estimates of the parameters when the simulated data have a normal error structure, except where the common support is limited in our example (that is, in the upper and lower tails of the distribution of the propensity score).

The parametric normal MTE estimator provides the best fit of the simulated data with a normal error structure (that is, the lowest sample standard deviations).[10] In contrast, it performs poorly in terms of bias when the error structure is nonnormal. Each of the other three MTE estimators are unbiased in this case; but the more semiparametric the method of estimation, the less precise the estimates tend to be.

To gauge the potential reliability of `margte` in real-world situations, we use the additional options of `margte_dgps` to modify the data-generating process to break certain identifying assumptions. To assess the impact of using an invalid instrument, we make the distance-to-college variable a direct function of the model's unobservables for one-third of the population. To test the effect of a limited common support for the propensity score, we reduce the magnitude of the coefficients on the observables in the treatment equation by 75%. Finally, we examine the result of reducing the sample size from 5,000 to 500 observations.

Figure 3 plots the sample mean of the MTE estimates for the parametric normal and semiparametric LIV estimators against their true values using our normally distributed simulated data with an invalid instrument. Neither MTE estimator produces an unbiased

10. Parametric normal MTE estimates calculated using the maximum likelihood routine of `movestay` are also unbiased with slightly smaller sample standard deviations.

estimate of the MTE in figure 3 except in a small region around the ATE (that is, $U_D = 0.5$). The MTE estimates are also now increasing in U_D, that is, decreasing in the propensity score, because of the positive correlation induced between the instrument `distCol` and u_0. The example highlights the crucial role played by the propensity score in the identification of the MTE so that its misspecification is a first-order concern regardless of whether parametric or semiparametric methods are used by `margte`.

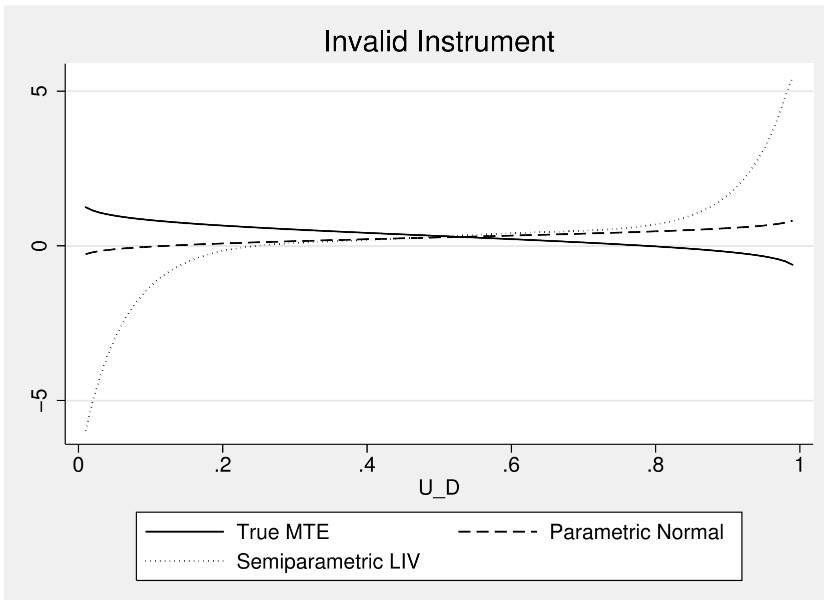

Figure 3. Estimating MTE with an invalid instrument

Figure 4 plots the sample mean of the MTE estimates of the parametric polynomial and semiparametric LIV estimators using our normally distributed simulated data with a limited common support for the propensity score. The range of values $[0.3, 0.7]$ for U_D with a well-defined common support in this example is typical of many empirical applications of MTE methods. Within this range, both estimators are unbiased. Outside this range, however, both suffer from a lack of identification, although the parametric polynomial estimator is not as imprecise because it retains some identification from its parametric restrictions.

Figure 4 also highlights the importance of examining the common support figure produced by `margte` whenever semiparametric estimation methods are used. Although positive frequencies of treated and untreated cases exist in this example at several values of U_D above 0.7 and below 0.3, they are often very small. This will have an effect on both the bias and the variance of the estimates. The same can be said for the variance of the estimates with small sample sizes, regardless of the estimator. Figure 5 plots 95% confidence intervals based on the sample standard deviation of the parametric normal MTE estimates in figure 3 with observation sizes of 500 and 5,000.

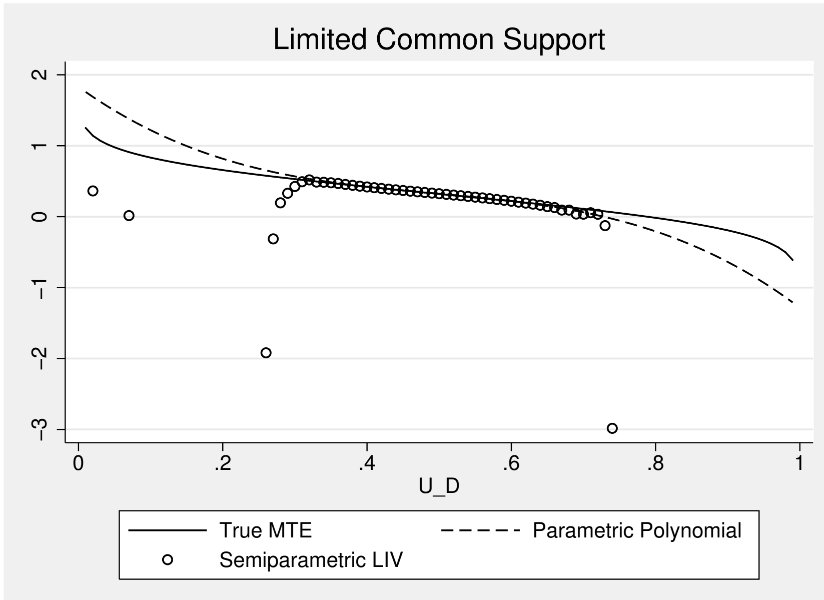

Figure 4. Estimating MTE with a limited common support

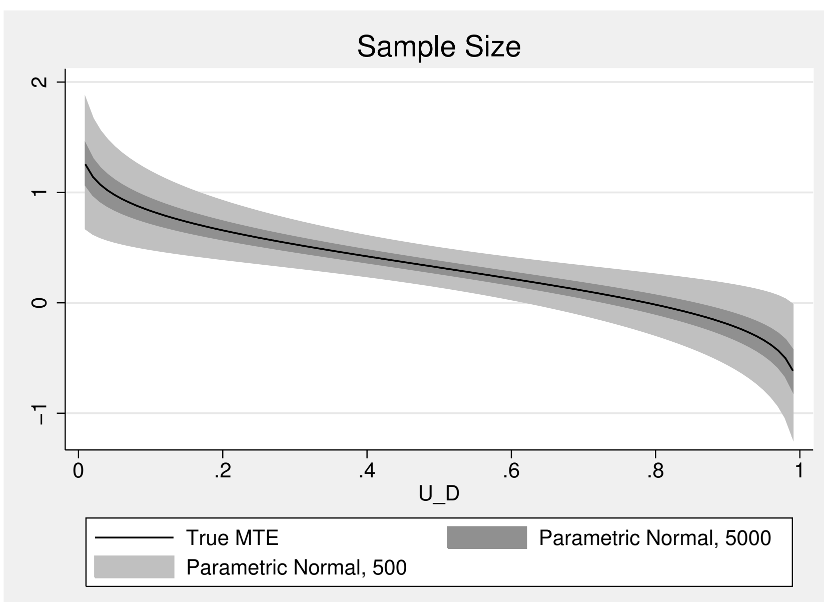

Figure 5. MTE precision and sample size

6 Summary

MTEs are quickly becoming a part of the toolkit for applied econometricians. The methods described here have been extended to include multinomial treatments and discrete instruments among other things.[11] They offer a convenient way to characterize the impact of a treatment that varies within a population in correlation with unobserved characteristics, and they serve as a reference point for conventional estimates of ATEs.

The `margte` command makes it possible to estimate both parametrically and semiparametrically ATEs and MTEs given a binary treatment and continuous outcome within the framework of the generalized Roy model. By nesting the existing Stata command `etregress` and the `movestay` command of Lokshin and Sajaia (2004) in its options, `margte` can also be used to produce two-step consistent and maximum likelihood estimates of traditional selection models for comparison and evaluation.[12]

Through a Monte Carlo simulation, we described the bias and variance properties of the four estimators supported by `margte`. All four produce unbiased estimates of ATE and MTE when the model for the propensity score is well specified and the range of values for the selection probability into treatment is well defined for both treated and untreated individuals. Parametric estimators do so with greater precision when the generalized Roy model is jointly normally distributed. When the joint distribution of the generalized Roy model is nonnormal, however, semiparametric estimators offer distinct advantages for both bias and variance.

7 References

Bjorklund, A., and R. Moffitt. 1987. The estimation of wage gains and welfare gains in self-selection models. *Review of Economics and Statistics* 69: 42–49.

Brinch, C. N., M. Mogstad, and M. Wiswall. 2012. Beyond LATE with a discrete instrument. Discussion Paper No. 703, Statistics Norway Research department.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil. 2011. Estimating marginal returns to education. *American Economic Review* 101: 2754–2781.

Doyle, J. J. 2007. Child protection and child outcomes: Measuring the effects of foster care. *American Economic Review* 97: 1583–1610.

Fan, J., and I. Gijbels. 1996. *Local Polynomial Modelling and Its Applications*. New York: Chapman & Hall/CRC.

French, E., and C. Taber. 2010. Identification of models of the labor market. In *Handbook of Labor Economics*, ed. O. Ashenfelter and D. Card, 537–617. Amsterdam: Elsevier.

11. See French and Taber (2010) for several examples in labor economics.
12. The accompanying help file for `margte` contains example syntax that can be used to replicate the output of the `etregress` and `movestay` commands.

Gutierrez, R. G., J. M. Linhart, and J. S. Pitblado. 2003. From the help desk: Local polynomial regression and Stata plugins. *Stata Journal* 3: 412–419.

Heckman, J. J. 2010. Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature* 48: 356–398.

Heckman, J. J., H. Ichimura, J. Smith, and P. Todd. 1998. Characterizing selection bias using experimental data. *Econometrica* 66: 1017–1098.

Heckman, J. J., D. Schmierer, and S. Urzua. 2010. Testing the correlated random coefficient model. *Journal of Econometrics* 158: 177–203.

Heckman, J. J., S. Urzua, and E. J. Vytlacil. 2006. Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics* 88: 389–432.

———. 2006b. Web supplement to understanding instrumental variables in models with essential heterogeneity: Estimation of treatment effects under essential heterogeneity. http://jenni.uchicago.edu/underiv/documentation_2006_03_20.pdf.

Heckman, J. J., and E. J. Vytlacil. 2001a. Local instrumental variables. In *Nonlinear Statistical Modeling: Proceedings of the Thirteenth Annual International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, ed. C. Hsiao, K. Morimune, and J. L. Powell, 1–46. New York: Cambridge University Press.

———. 2001b. Policy-relevant treatment effects. *American Economic Review* 91: 107–111.

Lokshin, M., and Z. Sajaia. 2004. Maximum likelihood estimation of endogenous switching regression models. *Stata Journal* 4: 282–289.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Marsh, C. 2006. locpolyslope. http://clmarsh.myweb.uga.edu.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

About the authors

Scott Brave is a senior business economist in the Economic Research Department of the Federal Reserve Bank of Chicago.

Thomas Walstrum is a senior associate economist in the Economic Research Department of the Federal Reserve Bank of Chicago and a PhD candidate in economics at the University of Illinois at Chicago.

A Appendix

Table 1 lists the actual values of the parameters of the generalized Roy model used to simulate the multivariate normally distributed data discussed in sections 4 and 5 by specifying the option `model(pnorm)` of `margte_dgps`. It also reports the sample mean and standard deviation (in parentheses) of the estimates of these parameters across all 5,000 random samples, as described in section 5.

Table 2 lists the actual values of the parameters of the generalized Roy model used to simulate the nonnormally distributed data in section 5 by specifying the option `dgp(poly)` of `margte_dgps`. It also reports the sample mean and standard deviation of the estimates of these parameters across all 5,000 random samples, as described in section 5.[13]

As an additional robustness check, we replicated the results in Heckman, Urzua, and Vytlacil (2006b) by using the `margte` command and the data that are available at http://jenni.uchicago.edu/underiv. The results of this exercise may be obtained from the authors upon request.

13. To compare the LIV estimator with the other polynomial estimators, we use `margte`'s `degree()` option to specify that a fourth-order polynomial expansion be used in the nonparametric regression of $\widetilde{Y}$ on $K(p)$.

Table 1. Actual and estimated parameters from simulated normal data

		Parametric		Semiparametric	
Parameter	Actual	Normal	Polynomial	LIV	Polynomial
$\beta_1^{\texttt{exp}}$	0.14	0.14	0.14	0.14	0.14
		(0.00)	(0.01)	(0.01)	(0.01)
$\beta_1^{\texttt{exp2}}$	−0.0012	−0.0012	−0.0012	−0.0012	−0.0012
		(0.0001)	(0.0004)	(0.0004)	(0.0004)
$\beta_1^{\texttt{momsEdu}}$	0.08	0.08	0.08	0.08	0.08
		(0.00)	(0.02)	(0.02)	(0.02)
ρ_1	−0.15	−0.15			
		(0.03)			
α_1	0.30	0.30			
		(0.07)			
$\beta_0^{\texttt{exp}}$	0.12	0.12	0.12	0.12	0.12
		(0.00)	(0.01)	(0.01)	(0.01)
$\beta_0^{\texttt{exp2}}$	−0.0015	−0.0015	−0.0015	−0.0015	−0.0015
		(0.0001)	(0.0001)	(0.0001)	(0.0001)
$\beta_0^{\texttt{momsEdu}}$	0.05	0.05	0.05	0.05	0.05
		(0.00)	(0.01)	(0.01)	(0.01)
ρ_0	0.25	0.25			
		(0.03)			
α_0	0.90	0.90			
		(0.06)			
ATE($\overline{x}$)	0.32	0.32	0.32	0.32	0.32
		(0.03)	(0.07)	(0.10)	(0.10)
MTE($\overline{x}, p = 0.1$)	0.83	0.83	0.84	0.84	0.84
		(0.06)	(0.34)	(0.44)	(0.43)
MTE($\overline{x}, p = 0.2$)	0.66	0.66	0.66	0.65	0.65
		(0.04)	(0.11)	(0.29)	(0.29)
MTE($\overline{x}, p = 0.3$)	0.53	0.53	0.53	0.53	0.53
		(0.03)	(0.11)	(0.16)	(0.16)
MTE($\overline{x}, p = 0.4$)	0.42	0.42	0.42	0.42	0.42
		(0.03)	(0.10)	(0.15)	(0.15)
MTE($\overline{x}, p = 0.5$)	0.32	0.32	0.32	0.32	0.32
		(0.03)	(0.07)	(0.14)	(0.14)
MTE($\overline{x}, p = 0.6$)	0.22	0.22	0.22	0.22	0.22
		(0.03)	(0.10)	(0.14)	(0.14)
MTE($\overline{x}, p = 0.7$)	0.11	0.11	0.11	0.11	0.11
		(0.03)	(0.11)	(0.16)	(0.16)
MTE($\overline{x}, p = 0.8$)	−0.02	−0.02	−0.02	−0.01	−0.01
		(0.04)	(0.11)	(0.28)	(0.28)
MTE($\overline{x}, p = 0.9$)	−0.19	−0.19	−0.20	−0.21	−0.20
		(0.06)	(0.33)	(0.42)	(0.41)

Table 2. Actual and estimated parameters from simulated polynomial data

Parameter	Actual	Normal	Parametric Polynomial	Semiparametric LIV	Semiparametric Polynomial
$\beta_{\texttt{exp}}$	0.12	0.13	0.12	0.12	0.12
		(0.00)	(0.01)	(0.01)	(0.01)
$\beta_{\texttt{exp2}}$	−0.0015	−0.0014	−0.0015	−0.0015	−0.0015
		(0.0001)	(0.0001)	(0.0001)	(0.0001)
$\beta_{\texttt{momsEdu}}$	0.05	0.07	0.05	0.05	0.05
		(0.01)	(0.01)	(0.01)	(0.01)
$\beta_{\texttt{expXp}}$	0.02	0.01	0.02	0.02	0.02
		(0.00)	(0.01)	(0.01)	(0.01)
$\beta_{\texttt{exp2Xp}}$	0.0003	0.0001	0.0003	0.0003	0.0003
		(0.0000)	(0.0002)	(0.0002)	(0.0002)
$\beta_{\texttt{momsEduXp}}$	0.03	−0.01	0.03	0.03	0.03
		(0.00)	(0.02)	(0.02)	(0.02)
β_p	3.40		3.40		3.40
			(0.76)		(0.76)
β_{p^2}	−12.00		−12.02		−12.02
			(3.07)		(3.07)
β_{p^3}	16.00		16.03		16.03
			(4.54)		(4.54)
β_{p^4}	−8.00		−8.01		−8.01
			(2.26)		(2.26)
α	0.90	0.84	0.90		0.90
		(0.07)	(0.11)		(0.11)
ATE($\overline{x}$)	0.32	0.32	0.32	0.32	0.32
		(0.03)	(0.07)	(0.11)	(0.11)
MTE($\overline{x}, p = 0.1$)	2.37	0.58	2.37	2.36	2.36
		(0.06)	(0.33)	(0.43)	(0.42)
MTE($\overline{x}, p = 0.2$)	1.18	0.49	1.18	1.18	1.18
		(0.05)	(0.12)	(0.28)	(0.28)
MTE($\overline{x}, p = 0.3$)	0.58	0.42	0.58	0.58	0.58
		(0.04)	(0.12)	(0.16)	(0.16)
MTE($\overline{x}, p = 0.4$)	0.35	0.37	0.35	0.35	0.35
		(0.03)	(0.10)	(0.14)	(0.14)
MTE($\overline{x}, p = 0.5$)	0.32	0.32	0.32	0.32	0.32
		(0.03)	(0.08)	(0.14)	(0.14)
MTE($\overline{x}, p = 0.6$)	0.29	0.27	0.29	0.29	0.29
		(0.03)	(0.10)	(0.14)	(0.14)
MTE($\overline{x}, p = 0.7$)	0.06	0.21	0.07	0.07	0.07
		(0.04)	(0.12)	(0.16)	(0.16)
MTE($\overline{x}, p = 0.8$)	−0.54	0.15	−0.54	−0.54	−0.54
		(0.05)	(0.12)	(0.28)	(0.28)
MTE($\overline{x}, p = 0.9$)	−1.73	0.06	−1.73	−1.73	−1.72
		(0.06)	(0.34)	(0.43)	(0.42)