User Guide of *MTE.exe*

James Heckman
University of Chicago and American Bar Foundation
Sergio Urzua
University of Chicago
Edward Vytlacil
Columbia University

March 22, 2006

The objective of this document is to provide a simple guide to the users of the code *MTE.exe*. This code implements different estimation procedures computing treatment parameters and IV estimates in the context of models with essential heterogeneity. See Heckman, Urzua, and Vytlacil (2006) for a discussion of models with essential heterogeneity.

1 The Two Ingredients

In order to use our code you need to consider two different files. The first file is our executable code. This code contains all the necessary routines for the estimation of the different treatment parameters and it is available at our website. The second file is a text file or also *init file*. This is a text file with the instructions to be used in the executable code.

2 How to Use the Files

The following are the practical steps that must be considered for the utilization of both the executable (exe) and *init* file.

Step 1. Click on the executable file

Step 2. A command prompt window will open. This window will display the following message:

name of the init file
(include path if not in the same folder as the exe file)

Step 3. You must write the name (and location if necessary) of your *init file*.

Step 4. If there are no problems with your init file, the estimation will begin. The code will automatically generate a diagnostic file with the information of the estimation procedure.

3 The Structure of the *init file*

The following is a line-by-line description of the init file. Please follow the order and make sure that the folders you are using in the *init* file exist. The code will display an error messages if there are problems in the init file.*

Line 1 (File Location\Name) -The location and name of the dataset.

Line 2 (Numeric) - The number of bootstraps to be used.

Line 3 (Numeric) - The number of variables in dataset.

Line 4 (Numeric) - The number of individuals in dataset.

*Please be aware that you need to fill all the lines with valid information. This means that even thought you are not using one of the estimation method, you need to give a valid name for the files.

Line 5 (Logical) - TRUE: We consider two different samples. The first one will be used to estimate the probit model. Since usually for the outcome equation we have missing values, the second sample is restricted to individual with non-missing values in the outcome equation. This is the sample used after the estimation of the propensity score. FALSE: We don't make the distinction and we do not observe missing values in outcomes.

Line 6 (Numeric) - The location of indicator that we will use to define the samples: 1 include without missing values in outcome, 0 otherwise.

Line 7 (Numeric) - The position (number of column) of the outcome variable in the dataset.

Line 8 (Numeric) - The number of X's in the outcome equation.

Line 9 (Numeric-Vector) - The position of the X's in dataset.

Line 10 (Numeric) - The position of the choice variable.

Line 11 (Numeric) - The number of Z's in the binary choice model.

Line 12 (Numeric-Vector) - The position of the Z's in dataset.

Line 13 (Numeric) - The number of instruments to be considered in the analysis.

Line 14 (Numeric-Vector or Scalar) - The position of the instruments in the data set.

Line 15 (Logical)l - TRUE: We use a parametric probit in the estimation of the choice model. FALSE: We do not use a parametric probit.

Line 16 (Logical) - TRUE: We use a non-parametric probit in the estimation of the choice model. FALSE: We do not use a nonparametric probit.

Line 17 (Logical) - TRUE: We use a linear probability model in the estimation of the choice model. FALSE: We do not use a linear probability model.[†]

Line 18 (Numeric) - Kernel Function: 0=Gaussian, 1= Epanechnikov, 2=Biweight.

Line 19 (Logical) - TRUE: We will estimate all the treatment parameters using only Parametric Normal Model. FALSE: We do not consider the parametric approach.

Line 20 (Logical) - TRUE: We will estimate all the treatment parameters using only Polynomials. FALSE: We do not consider the approximation of the control function by a polynomial in P.

Line 21 (Logical) - TRUE: We will estimate all the treatment parameters using only the semi-parametric approach 1 (LIV). FALSE: We do not consider the semiparametric approach 1 (LIV).

Line 22 (Logical) - TRUE: We will estimate all the treatment parameters using only the semi-parametric approach 2. FALSE: We do not consider the semiparametric approach 2.

Line 23 (Numeric) - If using a Polynomial of P, indicate the degree of the Polynomial.

Line 24 (File Location\Name) - This file contains the results of the estimated Propensity Score [ID, PPScore, D , INDICATOR (1 if included in support)].

Line 25 (File Location\Name) - This file contains the results for the density of the estimated propensity score by treatment groups[p, $\Pr(P = p|D = 0), \Pr(P = p|D = 1)$]. It can be used to generate a figure presenting the distribution of the propensity score by treatment status (see Figure 1 in the documentation file posted in our web site).

Line 26 (Logical) - TRUE: We consider directly the bootstrapping to get the confidence interval (CI). FALSE: 5% CI assuming normal distribution, i.e. we use the estimated standard deviation and the confidence level to construct the confidence interval using the critical values from a normal distribution.

Line 27 (Numeric) - The confidence interval level.

Line 28 (File Location\Name) - The location of file for the bootstrapping [Support, W_IV(First Instrument,....,Last Instrument), W_TT, W_TUT, W_ATE].

[†]One and only one of the arguments in Lines 15, 16 and 17 must be true.

Line 29 (File Location\Name) - The location of file for the bootstrapping [Support, MTE (estimated, mean, CI_L, CI_U)] from parametric approach.

Line 30 (File Location\Name) - The location of file for the bootstrapping [Support, MTE (estimated, mean, CI_L, CI_U)] from polynomial approach.

Line 31 (File Location\Name) - The location of file for the bootstrapping [Support, MTE (estimated, mean, CI_L, CI_U)] from semi-parametric approach 1 (LIV).

Line 32 (File Location\Name) - The location of file for the bootstrapping [Support, MTE (estimated, mean, CI_L, CI_U)] from semi-parametric approach 2.

Line 33 (File Location\Name) - The location of file for the BETA (parameters in the outcome equation) from parametric approach. The output is defined by the parameters we can recover in this case. See discussion in the text and our example.

Line 34 (File Location\Name) - The location of file for the BETA from polynomial approach. The output is defined by the parameters we can recover in this case. See discussion in the text and our example.

Line 35 (File Location\Name) - The location of file for the BETA from semi-parametric approach 1 (LIV). The output is defined by the parameters we can recover in this case. See discussion in the text and our example.

Line 36 (File Location\Name) - The location of file for the BETA from semi-parametric approach 2. The output is defined by the parameters we can recover in this case. See discussion in the text and our example.

Line 37 (File Location\Name) - The location of file for the GAMMA (parameters in the choice model). The file contains the parameters associated with the variables declared in line 12. Each line represents a variable and it has the following structure [estimated in original sample, mean from bootstrapping, CI_L, CI_U, stdev].

Line 38 (File Location\Name) - The location of file for the IV from parametric approach. Each line in this file contains the [estimated in original sample, mean from bootstrapping, stdev], second line is TUT[estimated, stdev]. The number of lines of the outcome is the number of instruments specified in line 11 plus one. This because the last line corresponds to the IV estimate when the propensity score is used as instrument.

Line 39 (File Location\Name) - The location of file for the IV from polynomial approach. The structure of the output is the same as in line 38.

Line 40 (File Location\Name) - The location of file for the IV from semiparametric approach 1 (LIV). The structure of the output is the same as in line 38.

Line 41 (File Location\Name) - The location of file for the IV from semiparametric approach 2. The structure of the output is the same as in line 38.

Line 42 (File Location\Name) - The location of file for the Treatment Parameters from parametric approach. First line is TT[estimated, mean from bootstrapping, stdev], second line is TUT[estimated, mean from bootstrapping, stdev], and third line is ATE[estimated, mean from bootstrapping, stdev].

Line 43 (File Location\Name) - The location of file for the Treatment Parameters from polynomial approach. The structure of the output is the same as in line 42.

Line 44 (File Location\Name) - The location of file for the Treatment Parameters from semiparametric approach 1 (LIV). The structure of the output is the same as in line 42.

Line 45 (File Location\Name) - The location of file for the Treatment Parameters from semiparametric approach 2. The structure of the output is the same as in line 42.

Line 46 (Numeric) - Bandwidth for the histogram of the propensity score. It is also used to define the full support.

Line 47 (Logical) - TRUE: We use optimal bandwidth to get each of the bandwidths for the first stage in the semi-parametric approach 1 (LIV) - crossvalidation. FALSE: We don't use crossvalidation.

Line 48 (Logical) - TRUE: We use optimal bandwidth to get each of the bandwidths for the first stage in the semi-parametric approach 1 (LIV) - Section 4.2 in Fan and Gijbels (1996). This alternative is very slow.[‡] FALSE: We don't use the procedure, and the bandwidth for each of the nonparametric regressions must be specified below.

[‡]Lines 46 and 47 cannot be true at the same time.

Line 49 (Numeric-Vector or Scalar) - If two previous lines are "FALSE" we need to specify the bandwidth. The bandwidth for the local linear regression the first stage in the semi-parametric approach 1. Therefore, if you are including n_X in the outcome equation (excluding the constant), you need to specify $2n_X + 1$ bandwidths. The last bandwidth is associated with the nonparametric regression of the dependent variable on the propensity score.

Line 50 (Logical) - TRUE: We use optimal bandwidth in the estimation of the MTE and LATE, using crossvalidation. FALSE: We do not use optimal bandwidth using crossvalidation.

Line 51 (Logical) - TRUE: We use optimal bandwidth in the estimation of the MTE and LATE, using Section 4.2 in Fan and Gijbels (1996). FALSE: We do not use optimal bandwidth, and a bandwidth must be specified below.

Line 52 (Numeric) - The Bandwidth for the MTE (if previous two FALSE)

Line 53 (File Location\Name) - The location of the file where we save the information of the bandwidth selection.

Line 54 (Logical) - TRUE: We use a discrete approximation of the derivative. FALSE: We use the nonparametric estimator of the derivative.

Line 55 (Logical) - TRUE: We consider a local **linear** regression to get the MTE. FALSE: We do not use a local linear approximation and the degree of the local polynomial must be provided in Line 56.

Line 56 (Numeric) - If above (Line 55) FALSE, we need to specify the order of the polynomial for the local approximation.

Line 57 (Logical) - TRUE: You ask to save the common support data. FALSE: You do not save the data

Line 58 (File Location\Name) - In this file you obtain the data set used in the estimation of the marginal treatment effect. This sample includes only observations in the common support of the propensity score.

Line 59 (Logical) - FALSE: NO LATEs considered. TRUE: LATEs are going to be estimated.

Line 60 (Numeric) - Number of LATE estimators to be considered

Line 61 (Numeric-Vector) - Pairs of (p_2, p_1) to be considered in the analysis where $p_2 > p_1$

Line 62 (File Location\Name) - The location of file for the LATE(estimated in original sample, mean from bootstrapping, CI_L, CI_U, stdev) - parametric

Line 63 (File Location\Name) - The location of file for the LATE(estimated in original sample, mean from bootstrapping, CI_L, CI_U, stdev) - polynomial

Line 64 (File Location\Name) - The location of file for the LATE(estimated in original sample, mean from bootstrapping, CI_L, CI_U, stdev)- semiparametric 1 (LIV)

Line 65 (File Location\Name) - The location of file for the LATE(estimated in original sample, mean from bootstrapping, CI_L, CI_U, stdev)- semiparametric 2

The following is an example of the *init* file. It uses the data set `data_5000.out` available at http://jenni.uchicago.edu/underiv/. The code *MTE exe* is also available at this website.

References

Heckman, J. J., S. Urzua, and E. J. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics*.

```
C:\data_5000.out                                    ! Line 1
50                                                  ! Line 2
7                                                   ! Line 3
5000                                                ! Line 4
false                                               ! Line 5
1                                                   ! Line 6
2                                                   ! Line 7
3                                                   ! Line 8
1 4 5                                               ! Line 9
3                                                   ! Line 10
3                                                   ! Line 11
1 6 7                                               ! Line 12
2                                                   ! Line 13
6 7                                                 ! Line 14
true                                                ! Line 15
false                                               ! Line 16
false                                               ! Line 17
2                                                   ! Line 18
true                                                ! Line 19
true                                                ! Line 20
true                                                ! Line 21
true                                                ! Line 22
4                                                   ! Line 23
C:\PPscore.out                                      ! Line 24
C:\\KernelPPScore.out                               ! Line 25
false                                               ! Line 26
5                                                   ! Line 27
C:\Weigths.out                                      ! Line 28
C:\Parametric\MTE_boots.out                         ! Line 29
C:\Polynomial\MTE_boots.out                         ! Line 30
C:\Semiparametric1\MTE_boots.out                    ! Line 31
C:\Semiparametric2\MTE_boots.out                    ! Line 32
C:\Parametric\BETA.out                              ! Line 33
C:\Polynomial\BETA.out                              ! Line 34
C:\Semiparametric1\BETA.out                         ! Line 35
C:\Semiparametric2\BETA.out                         ! Line 36
C:\GAMMA.out                                        ! Line 37
C:\Parametric\IV.out                                ! Line 38
C:\Polynomial\IV.out                                ! Line 39
C:\Semiparametric1\IV.out                           ! Line 40
C:\Semiparametric2\IV.out                           ! Line 41
C:\Parametric\TREATMENT.out                         ! Line 42
C:\Polynomial\TREATMENT.out                         ! Line 43
C:\Semiparametric1\TREATMENT.out                    ! Line 44
C:\Semiparametric2\TREATMENT.out                    ! Line 45
0.01                                                ! Line 46
true                                                ! Line 47
false                                               ! Line 48
0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3             ! Line 49
false                                               ! Line 50
false                                               ! Line 51
0.3                                                 ! Line 52
C:\bandwidth.out                                    ! Line 53
false                                               ! Line 54
true                                                ! Line 55
2                                                   ! Line 56
true                                                ! Line 57
C:\commonsuptdata.out                               ! Line 58
true                                                ! Line 59
3                                                   ! Line 60
0.38d0 0.62d0 0.55d0 0.79d0 0.21d0 0.45d0           ! Line 61
C:\Parametric\LATE.out                              ! Line 62
C:\Polynomial\LATE.out                              ! Line 63
```

```
C:\Semiparametric1\LATE.out                    ! Line 64
C:\Semiparametric2\LATE.out                    ! Line 65
```