11/23/2017

Prepared by Xintong Wang; edited by Carlos Flores

Description of files needed to obtain the bounds in Flores and Flores-Lagunes (2010, 2013).

The folder "To_Share_in_IADB_MATE" consists of the following do-files and data set.

1. **MATE_InvalidIV.do. This is the main file** where the variables are defined. Under the assumptions in Flores and Flores-Lagunes (2010, 2013), this program: (i) estimates the point identified quantities and testable implications discussed in those papers; (ii) estimates the bounds on LNATE_at, LNATE_nt, MATE, NATE, LATE derived in those papers, and calculates the corresponding 95% Confidence Intervals. It calls the two other do files, A1.do and MATE_InvalidIV_Bounds.do. This do file starts the bootstrap and then calls the do-file MATE_InvalidIV_Bounds.do to use the Chernozhukov, Lee and Rosen (2013) methodology to compute the estimated bounds and 95% Confidence Intervals. The definitions of the variables are as follows:

   "Outcome" is the outcome variable.

   "treatment" is the Randomized treatment (or the invalid IV).

   "S" is the mechanism (or the endogenous treatment).

   "W" is the sampling weight.

   "mprid" is the id number (identifier) that breaks the tie of the quantiles when calculating trimming bounds.

2. **A1.do** calculates the bounds (or bounding functions if *min* or *max* operators are present) on LNATE_at, LNATE_nt, MATE, NATE, LATE under the assumptions in Flores and Flores-Lagunes (2010, 2013). This do file is called by the program MATE_InvalidIV.do.

   Here is a brief explanation of all the arguments in A1.do:
   y is the outcome
   z is the Randomized treatment (or the invalid IV).
   s is the mechanism (or the endogenous treatment).
   wgt is the sampling weight.
   sv is the id number that breaks the tie of the quantiles when calculating the trimmed bounds.

3. **MATE_InvalidIV_Bounds.do**. This do file implements the Chernozhukov, Lee and Rosen (2013) methodology to compute the estimated bounds and 95% Confidence Intervals. It also summarizes all the estimation results in the two tables below, which appear at the end of the log file. Example:

Table 1. Point Estimates of Interest

Means[25,2]

	Mean	Standard E~s
# of obser~s	8020	0
ITT	22.217732	4.8199062
Compliers	.21001454	.01132356
LATE	105.7914	22.995847
Always-tak~s	.44653124	.00919192
Never-takers	.34345422	.00650787
pr_at_c	.65654578	.00650787
pr_nt_c	.55346876	.00919192
Y11	213.44627	3.7919159
Y00	144.24089	4.2250962
Y01	195.05979	5.9101954
Y10	142.70782	4.6401472
EY[Z1]	189.15085	3.0294675
EY[Z0]	166.93312	3.604243
EY[D1]	206.32529	3.3142262
EY[D0]	143.6271	3.1004994
Imp_a	50.818898	7.2208489
Imp_b	-70.738442	5.929529
Imp_c	69.205375	5.902312
Min y	0	0
Max y	2358.4182	151.77971
PrZ0	.48170368	.00577355
PrZ1	.51829632	.00577355
PrS0	.444619	.00575981
PrS1	.555381	.00575981

Each line in the above table is as follows.

Notation: $Z = 0,1$ is the randomized treatment (or the invalid IV); $D = 0,1$ is the mechanism (or the endogenous treatment); and Y is the outcome.

(1) # of obser~s is the number of observations
(2) ITT $E[Y|Z = 1] - E[Y|Z = 0]$
(3) Compliers $\pi_c = Pr[D = 1|Z = 1] - Pr[D = 1|Z = 0]$
(4) LATE (the IV estimates) $(E[Y|Z = 1] - E[Y|Z = 0])/(Pr[D = 1|Z = 1] - Pr[D = 1|Z = 0])$
(5) Always-takers $\pi_{at} = Pr[D = 1|Z = 0]$
(6) Never-takers $\pi_{nt} = Pr[D = 0|Z = 1]$
(7) Pr_at_c $Pr[D = 1|Z = 1]$
(8) Pr_nt_c $Pr[D = 0|Z = 0]$
(9) Y11 $E[Y|Z = 1, D = 1]$
(10) Y00 $E[Y|Z = 0, D = 0]$
(11) Y01 $E[Y|Z = 0, D = 1]$

(12) Y10 $E[Y|Z=1, D=0]$
(13) EY[Z1] $E[Y|Z=1]$
(14) EY[Z0] $E[Y|Z=0]$
(15) EY[D1] $E[Y|D=1]$
(16) EY[D0] $E[Y|D=0]$
(17) Imp_a. Testable Implication: $E[Y|Z=0, D=1] - E[Y|Z=0, D=0]$
(18) Imp_b. Testable Implication: $E[Y|Z=1, D=0] - E[Y|Z=1, D=1]$
(19) Imp_c. Testable Implication: $E[Y|Z=1, D=1] - E[Y|Z=0, D=0]$
(20) Miny: $\min\{Y\}$
(21) Maxy: $\max\{Y\}$
(22) $\Pr[Z=0]$
(23) $\Pr[Z=1]$
(24) $\Pr[D=0]$
(25) $\Pr[D=1]$

Table 2. Bounds Estimates

	Ass1&3: Lower bound	Ass1&3: Upper bound	Ass1&4: Lower bound	Ass1&4: Upper bound	Ass1&5: Lower bound	Ass1&5: Upper bound	Ass1,4&5: Lower bound	Ass1,4&5: Upper bound
Lnate_nt	-89.510267	114.44017	0	114.44017	-1.508818	114.44017	-9.188e-19	114.44017
CI_Lnate~95%	-106.69064	125.99206	0	126.12963	-12.327258	126.1205	-3.748e-16	126.12963
Lnate_at	-96.105107	118.17	0	118.17	18.412884	118.17	18.35783	118.17
CI_Lnate~95%	-109.33737	135.44491	0	136.1699	6.1367931	136.4112	6.1892625	136.41083
Lnate_c	.	.	0	456.59366	-52.372255	168.99095	0	69.182861
CI_Lnate~95%	.	.	0	474.25554	-65.116689	185.31176	0	79.493283
Lmate_c	.	.	0	456.59366	-170.60418	70.715825	0	71.714556
CI_Lmate~95%	.	.	0	474.25554	-186.7156	80.607308	0	80.970177
MATE	.	.	0	23.58687	-36.246706	14.850288	0	15.808665
CI_MATE_95%	.	.	0	32.057648	-42.452254	17.477709	0	18.06509
NATE	.	.	5.684e-11	22.199347	7.3801873	58.521119	5.5165465	22.199347
CI_NATE_95%	.	.	-2.765e-08	31.160492	-1.3925598	70.310327	.74253351	31.295311
LATE	.	.	0	111.53983	-170.60418	70.715825	0	74.56312
CI_LATE_95%	.	.	0	152.47511	-186.7156	80.607308	0	83.11981

The bounds in Table 2 are defined as follows.

Lnate_nt: $E[Y(1, D(0))|nt] - E[Y(0, D(0))|nt]$ (note: since $D(1) = D(0)$ for never-takers, Lnate_nt also equals $E[Y(1, D(1))|nt] - E[Y(0, D(0))|nt]$ and $E[Y(1, D(1))|nt] - E[Y(0, D(1))|nt]$. The same holds true for the always-takers.)

Lnate_at: $E[Y(1, D(0))|at] - E[Y(0, D(0))|at]$

Lnate_c: $E[Y(1, D(0))|c] - E[Y(0, D(0))|c]$

Lmate_c: $E[Y(1, D(1))|c] - E[Y(1, D(0))|c]$

MATE: $E[Y(1, D(1)) - Y(1, D(0))]$

NATE: $E[Y(1, D(0)) - Y(0, D(0))]$

LATE: $LATE = \frac{MATE}{E[D(1) - D(0)]}$

The assumption numbers used in Table 2 correspond to those used in Flores and Flores-Lagunes (2013) in the context of bounding the Local Average Treatment effect (on compliers) with an Invalid Instrument (see paper for details):

Assumption 1: Randomly Assigned Instrument

Assumption 2: Nonzero Average Effect of the Instrument on the Endogenous Treatment

Assumption 3: (Positive) Individual-Level Monotonicity of the Instrument on the Endogenous Treatment.

Assumption 4: Weak (Positive) Monotonicity of Mean Potential Outcomes and Counterfactual Outcomes *within* Strata

Assumption 5: Weak Monotonicity of Mean Potential Outcomes and Counterfactual Outcomes *across* Strata (where the always takers have the largest average potential/counterfactual outcomes, and the never takers have the lowest)

4. **CLR_01272015.ado**. This ado file implements the Chernozhukov, Lee and Rosen (2013) methodology to estimate the bounds and obtain confidence intervals. It is called by the do-file MATE_InvalidIV_Bounds.do.
5. **mcholx_01272015.ado**. This ado file is use to solve singularity issues of the variance covariance matrix that may arise while implementing the Chernozhukov, Lee and Rosen (2013) methodology.
6. **JC_Data_IADB_MATE.dta**. This is the data set to be used as an example. It is based on the one used in Flores and Flores-Lagunes (2010, 2013).