MANSKI AND LEE BOUNDS

Methods for accounting for attrition vary according to whether we assume that attrition is unsystematic between the treatment and control groups or not. There are several standard techniques to deal with attrition and sample selection; I focus in this post on how to code boundary approaches proposed by Manski (1990) and Lee (2009)[1].

To illustrate the method, I use data from a randomized controlled trial (RCT) conducted by Blattman et al. (2011, 2014). The data can be downloaded from Harvard Dataverse[2].

Suppose we are interested in investigating the impact of randomly assigned cash transfers (given at baseline) on migration (measured at endline). After examining the data, we find that there is a high level of attrition at endline (25% at the first endline) and that this might affect the validity of our results because it is likely that attrition and migration are related, in perhaps a non-systematic way. For example, what if the attrited themselves are the people who migrated?

The first boundary approach (Manski Upper Bound) proposes an extreme assumption that all the attriters in the treatment group are migrants and all the attriters in the control group are not migrants. This generates an upper bound. The lower bound would thus be the exact opposite: all the attriters in the treatment group are non-migrants and all the attriters in the control group are migrants.

Figure 1.a illustrates how to construct these bounds in Stata[3], where the bounds themselves then become the two extreme cases for the outcome under study.

Figure 1- STATA Code for Manski Bounds and Lee Bounds
1.a General Code

```
********************************************************
************************1. MANSKI BOUNDS****************
********************************************************

***(1)Generate Higher Bound
gen manski_upperbound=outcome
quietly sum manski_upperbound
replace manski_upperbound = r(max) ///
if manski_upperbound ==. & treatment ==1
replace manski_upperbound = r(min) ///
if manski_upperbound ==. & treatment ==0

***(2)Generate Lower Bound
gen manski_lowerbound=outcome
quietly sum manski_lowerbound
replace manski_lowerbound = r(min) ///
if manski_lowerbound ==. & treatment ==1
replace manski_lowerbound = r(max) ///
if manski_lowerbound ==. & treatment ==0


********************************************************
************************2. LEE BOUNDS*******************
********************************************************

leebounds migrate_e assigned
```

---

[1] Other boundary methods include that suggested by Behaghel et al. (2012) where the bounds are constructed based on the number of times it takes to reach a respondent, hence trimming the "extra" respondents above a threshold (identified by the intensity of the effort it gets to reach a respondent) in either of the treatment or control groups.
[2] I am using the dataset "yop2_yop4_deid.dta".
[3] The Stata version used here is Stata 17.0.
[4] We are looking at the results for the first endline only for simplicity.

Figure 1.b shows the code for our specific example, where the outcome is migration. Columns 1 and 2 of table 1 display the results4.

1.b Example-Specific Code

```
use "yop2_yop4_deid.dta"
keep if endline ==1

******************************************************************
*************************1. MANSKI BOUNDS*************************
******************************************************************

***(1)Generate Higher Bound
gen manski_upperbound=migrate_e
replace manski_upperbound = 1 ///
if manski_upperbound ==. & assigned ==1
replace manski_upperbound = 0 ///
if manski_upperbound ==. & assigned ==0

***(2)Generate Lower Bound
gen manski_lowerbound=migrate_e
replace manski_lowerbound = 0 ///
if manski_lowerbound ==. & assigned ==1
replace manski_lowerbound = 1 ///
if manski_lowerbound ==. & assigned ==0

***Run the regression***
reg manski_upperbound assigned
outreg2 using bounds.doc, replace ctitle(Manski Upper Bound) se r2
reg manski_lowerbound assigned
outreg2 using bounds.doc, append ctitle(Manski Lower Bound) se r2

******************************************************************
*************************2. LEE BOUNDS*************************
******************************************************************

leebounds migrate_e assigned
outreg2 using bounds.doc, append ctitle(Lee Bound) se r2
```

Note that an underlying assumption of the Manski bound is that the outcome variable should be bounded. The second approach discussed below does not need this to be the case.

The second boundary approach Lee Bounds is a type of trimming procedure, where the group (either control or treatment) that potentially suffers the most from sample attrition is trimmed (Tauchmann, 2014). The lower and upper bound are computed by making assumptions about the missing data of the outcomes of the attrition group. Lee (2009) bounds rely on a few assumptions only, namely random assignment of treatment and monotonicity, where the latter means that the treatment status affects attrition in just one direction. In other words, being assigned to treatment makes attrition either more or less likely for any observation. There is a simple built-in command leebounds in Stata that takes care of the process. But what is leebounds exactly doing?

1. It first calculates the below ratio:

$$x = \frac{fraction\ remaining\ in\ less\ attrited\ group - fraction\ remaining\ in\ more\ attrited\ group}{fraction\ remaining\ in\ less\ attrited\ group}$$

2. It then drops the lowest percentages of the outcomes from the less attrited group.
3. Afterwards, it creates the first bound by calculating the mean outcome for the trimmed group (less attrited after dropping the lowest percentages), which is then compared to the mean outcomes in the group with lower attrits.
4. To create the lower bound, it drops the highest percentages of the outcomes from the less attrited group.
5. Lastly, it calculates the mean outcome for the trimmed group (less attrited after dropping the highest percentages), which is then compared to the mean outcomes in the group with lower attrits.

Figure 1.a displays the simple code and Figure 1.b shows the code for our specific example. Columns 1 and 2 of table 1 display the results. The series of figures below show the detailed commands in Figures 1.a and 1.b being implemented, followed by running the relevant regressions[5].

```
. keep if endline==1
(2,677 observations deleted)

. do "/var/folders/py/jdhh556n7tqb1bzp3dmsfz880000gn/T//SD02003.000000"


. ********************************************************************
. *************************1. MANSKI BOUNDS**************************
. ********************************************************************
.
. ***(1)Generate Higher Bound
. gen manski_upperbound=migrate_e
(79 missing values generated)

. replace manski_upperbound = 1 ///
> if manski_upperbound ==. & assigned ==1
(8 real changes made)

. replace manski_upperbound = 0 ///
> if manski_upperbound ==. & assigned ==0
(71 real changes made)


.
. ***(2)Generate Lower Bound
. gen manski_lowerbound=migrate_e
(79 missing values generated)

. replace manski_lowerbound = 0 ///
> if manski_lowerbound ==. & assigned ==1
(8 real changes made)

. replace manski_lowerbound = 1 ///
> if manski_lowerbound ==. & assigned ==0
(71 real changes made)
```

```
.
. ***Run the regression***
. reg manski_upperbound assigned

      Source |       SS           df       MS      Number of obs   =     2,677
-------------+----------------------------------   F(1, 2675)      =     15.01
       Model |  3.29019259         1  3.29019259   Prob > F        =    0.0001
    Residual |  586.399759     2,675  .219214863   R-squared       =    0.0056
-------------+----------------------------------   Adj R-squared   =    0.0052
       Total |  589.689951     2,676  .220362463   Root MSE        =    .4682

------------------------------------------------------------------------------
manski_upp~d | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    assigned |   .0701195   .0180993     3.87   0.000     .0346293    .1056096
       _cons |   .2928994   .0127335    23.00   0.000      .267931    .3178678
------------------------------------------------------------------------------

. outreg2 using bounds.doc, replace ctitle(Manski Upper Bound) se r2
bounds.doc
dir : seeout
```

[5] It is worth noting that there exists a more general Stata command called "tebounds" that applies various methods to bound the average treatment effect when the treatment and the outcome are binary variables and when the treatment assignment is endogenous and misreported (McCarthy, Millimet, and Roy 2015).

```
. reg manski_lowerbound assigned

      Source |       SS           df       MS      Number of obs   =     2,677
-------------+----------------------------------   F(1, 2675)      =      0.39
       Model |  .089532452         1  .089532452   Prob > F        =    0.5309
    Residual |  609.839493     2,675  .22797738    R-squared       =    0.0001
-------------+----------------------------------   Adj R-squared   =   -0.0002
       Total |  609.929025     2,676  .227925645   Root MSE        =    .47747

------------------------------------------------------------------------------
manski_low~d | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    assigned |    .0115669   .0184575     0.63   0.531    -.0246256    .0477594
       _cons |    .3454142   .0129855    26.60   0.000     .3199516    .3708768
------------------------------------------------------------------------------

. outreg2 using bounds.doc, append ctitle(Manski Lower Bound) se r2
bounds.doc
dir : seeout
```

```
. ********************************************************************
. *************************2. LEE BOUNDS*****************************
. ********************************************************************
.
. leebounds migrate_e assigned

Lee (2009) treatment effect bounds

Number of obs.                    =    2677
Number of selected obs.           =    2598
Trimming porportion               =  0.0468


------------------------------------------------------------------------------
   migrate_e | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
assigned     |
       lower |    .0185804   .0194879     0.95   0.340    -.0196152    .0567761
       upper |    .0676335    .019122     3.54   0.000     .0301551    .1051118
------------------------------------------------------------------------------

. outreg2 using bounds.doc, append ctitle(Lee Bound) se r2
bounds.doc
dir : seeout
```

Table 1- Manski Bounds and Lee Bounds: Impact of cash transfers on migration

VARIABLES	(1) Migration Manski Upper Bound	(2) Migration Manski Lower Bound	(3) Migration Lee Bound
Assigned Treatment Status	0.0701***	0.0116	
	(0.0181)	(0.0185)	
Lee Bounds (trimming highest ~5% of observations)			0.0186
			(0.0195)
Lee Bounds (trimming lowest ~5% of observations)			0.0676***
			(0.0191)
Constant	0.293***	0.345***	
	(0.0127)	(0.0130)	
Observations	2,677	2,677	2,677
R-squared	0.006	0.000	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

References

Behaghel, Luc, Bruno Crepon, Marc Gurgand, and Thomas Le Barbanchon. 2012. "Please Call Again: Correcting Non-Response Bias in Treatment Effect Models." SSRN Journal. https://doi.org/10.2139/ssrn.2157893.

Blattman, Christopher; Fiala, Nathan; Martinez, Sebastian, 2014, "Northern Uganda Social Action Fund - Youth Opportunities Program", https://doi.org/10.7910/DVN/27898, Harvard Dataverse, V1.

Blattman, Christopher; Fiala, Nathan; Martinez, Sebastian, 2019, "The long term impacts of grants on poverty: 9-year evidence from Uganda's Youth Opportunities Program", https://doi.org/10.7910/DVN/V0N0HA Harvard Dataverse, V1.

Duru, M., & Kopper, S. (n.d.). Data Analysis. The Abdul Latif Jameel Poverty Action Lab (J-PAL). Retrieved January 27, 2022, from https://www.povertyactionlab.org/resource/data-analysis.

Lee, David. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." The Review of Economic Studies 76 (3): 1071–1102. 10.3386/w11721

McCarthy, Ian, Daniel L. Millimet, and Manan Roy. 2015. "Bounding Treatment Effects: A Command for the Partial Identification of the Average Treatment Effect with Endogenous and Misreported Treatment Assignment." The Stata Journal 15 (2): 411–36. https://doi.org/10.1177/1536867x1501500205.

Tauchmann, Harald. 2014. "Lee (2009) Treatment-Effect Bounds for Nonrandom Sample Selection." The Stata Journal 14 (4): 884–94. 10.1177/1536867x1401400411.

Razan Amine, Research Assistant in Economics
16 February 2022