Paper: Crime and Mismeasured Punishment: Marginal Treatment Effect with Misclassification

Author: Vitor Possebom

Journal: Review of Economics and Statistics

Replication Package

1. Empirical Application

The R Project *possebom_mte_misclassification_replication_package.Rproj* allows you to run the entire empirical application (Section 5 and Appendix J) without the need to change your working directory.

The dataset *d_small_crimes_complete.rds* in the folder *datasets* is the raw data file. Its variables are described below.

1. defendant: Name of the defendant
2. id_process: Unique identifying number of the criminal case in the São Paulo's Court System
3. allocation_date: date when the case was allocated to a judge
4. allocation_year: year when the case allocated to a judge
5. cjpg_date_disp: date when the trial judge wrote its ruling
6. final_date: date of the final ruling (either by the trial judge or the Appeals Court) of the case
7. juiz: name of the randomly allocated trial judge
8. cjpg_comarca: administrative unit of the case
9. cjpg_foro: court district of the case. Trial judges are randomly allocated at this geographic level.
10. cjpg_vara: courtroom of the case
11. cjpg_magistrado: name of the judge who wrote the trial court's ruling
12. cposg_distribuicao: Appeals Court's district
13. cposg_relator: Appeals Court's Judge Rapporteur
14. cposg_revisor: Appeals Court's judge who revises the case
15. cposg_composicao: Appeals Court's composition
16. cposg_decisoes: Appeals Court's rulings
17. dismissed: whether the case was dismissed by the trial judge
18. NPA: whether the defendant signed a non-prosecution agreement
19. convicted: whether the trial judge convicted the defendant

20. punished: whether the defendant was punished (= NPA + convicted)
21. appealed: whether the case went to the Appeals Court
22. overruled: whether the Appeals Court reversed the decision of the trial judge
23. final_punished: whether the defendant was punished according to the final ruling in her case
24. reoffendend: whether the defendant recidivated at some point during my sampling period
25. date_reoffense: date of the recidivism event
26. duration: number of days between the case's final ruling's date and the recidivism event

Starting from the dataset *d_small_crimes_complete.rds*, all results are built using the R scripts whose names start with *codeXX*.

1. *code01_defining-the-instrument.R*: It creates the leave-one-out punishment rate that is used as an instrument in the empirical application. It produces Figure I.1a
2. *code02_defining-the-outcome-variable.R*: It creates the outcome variable (recidivism event within two years of the final rulings' date) of the empirical application.
3. *code03_districts_with_more_than_2_judges.R*: It selects court districts with more than two judges during my sampling period.
4. *code04_descriptive_stats_misclassification.R*: It analyzes whether a treatment variable defined based on the trial judges' decision misclassifies the true treatment defined based on final rulings. It produces Figure 1 and Table J.1.
5. *code05_descriptive_stats_outcome.R*: It analyzes the unconditional and conditional means of the outcome variable. It produces Table I.1.
6. *code06_descriptive_stats_instrument.R*: It analyzes the distribution of the instrumental variable. It produces Figure I.1b.
7. *code07_mte-results-correctly-measured.R*: It estimates the parametric MTE function using the correctly measured treatment variable. It produces the first two columns of Table J.2.
8. *code08_mte-results-misclassified.R*: It estimates the parametric MTE function using the misclassified treatment variable. It produces the last two columns of Table J.2.
9. *code09a_mte-bounds-one-value-of-c.R*: It produces Figure 3.a and Figure 3.b.
10. *code09b_mte-bounds-one-value-of-c-with-ci.R*: It produces Figure J.2b.
11. *code09c_mte-bounds-two-values-of-c.R*: It produces Figure J.3.
12. *code09d_mte-bounds-with-ci-for-the-mte.R*: It produces Figure J.2a.

13. *code10_mte-sign.R*: It produces Figure J.1.
14. *code11_mte-contained-by-bounds.R*: It finds that the estimated bounds contain the entire MTE function for every court district.
15. *code12a_analyzing-LIV-bias.R*: It produces Table J.3's Panel A.
16. *code12b_analyzing-LIV-bias_trimmed.R*: It produces Table J.3's Panel B.
17. *code14_standard-2sls.R*: It produces Table J.4.
18. *code15_standard-mte.R*: It produces Figure J.5.
19. *code17_covariate-balance.R*: It produces the results described in the last paragraph of Appendix I.1.
20. *code18_semiparametric-propensity-scores.R*: It produces Figure J.4.
21. *code19_number-of-words-per-name.R*: It counts the number of words in the defendant's name.
22. *code20_compare_ps_derivatives.R*: It produces Figure 2.a and Figure 2.b.
23. *code21_average-mte-function.R*: It estimates the average MTE function and LIV estimand across court districts. These results are not included in the paper.
24. *code22_find_liv_within_minus1_and_1.R*: It checks if there is any court district whose entire LIV estimand lies within -1 and 1.
25. *function_liv_estimand.R*: It is an auxiliary code that estimates the LIV estimand.
26. *function_standard-mte.R*: It is an auxiliary code that estimates the semiparametric MTE function.

All results were produced using R version 4.1.1 (2021-08-10) in a Microsoft Windows Version 22H2 (OS Build 22621.1992).

2. Monte Carlo Simulation

The R Project *monte_carlo.Rproj* allows you to run the entire Monte Carlo Simulation (Appendix H.4) without the need to change your working directory.

1. R script *code01_monte_carlo_simulation.R* is the main code file. It runs the entire simulation, producing Figure H.1a (*figure_MC_bias_MTE.pdf*), Figure H.1b (*figure_MC_bias_LIV.pdf*), Figure H.2a (*figure_MC_bounds_contain_MTE_X0.pdf*), Figure H.2b (*figure_MC_bounds_contain_MTE_X1.pdf*) and Figure H.2c (*figure_MC_bounds_contain_MTE_X2.pdf*).
2. R script *function_liv_estimand.R* is an auxiliary code that estimates the LIV estimand.

All results were produced using R version 4.1.1 (2021-08-10) in a Microsoft Windows Version 22H2 (OS Build 22621.1992).