Replication Files

This file summarizes the programs in MATLAB and Stata which reproduce the results in the paper: “Assessing Sensitivity to Unconfoundedness: Estimation and Inference” by Matthew Masten, Alexandre Poirier and Linqi Zhang.

MATLAB programs

There are two main parts contained in the codeMatlab folder:

1) Programs that reproduce the Monte Carlo results in Appendix H.

2) Programs that reproduce the main sensitivity analysis on the LaLonde1986 dataset.

Monte Carlo

1- Impact of functional form misspecification: Appendix H.1, Table S1, Figure S4

2- Choice of tuning parameters: Appendix H.2, Figure S5

Empirical Illustration

It refers to section 6 of the paper -- sensitivity analysis illustration using data on the National Supported Work demonstration project studied by Lalonde (1986). Our analysis uses two different versions of the LaLonde dataset, as analyzed by Dehejia and Wahba (1999): the experimental sample (n=445), and the observational sample (n=390). See section 6 for more description of the data construction. The MATLAB program outputs estimated treatment effect bounds under c-dependence, along with corresponding pointwise confidence bands.

We provide all the code necessary to go from the raw data to the results of the paper. Run the script `main_MC_misspec.m` to reproduce the results in Appendix H.1. Run the script `main_MC_coverageProb.m` to reproduce the results in Appendix H.2. Run the script `main_Empirial.m` to reproduce Figure 1 in section 6. The programs are independent of each other, so each should be run without a problem as long as all necessary subfolders are present. Details of the main programs are described below.

The programs involve MATLAB parallel computing in several places. It is important to change the code that opens the parallel pool, so that it is compatible with the configuration of your operating system. By default, the code that opens the parallel pool is commented out.

Detailed description of main MATLAB programs

1. `main_Empirical.m`

Description

- Confidence sets estimation for bounds on ATE and ATT for the lalonde1986 application.

Inputs

- lalonde1986.dat
- tuningChoiceVec: tuning parameter index
- alphaVec: significance level

Main output files

- images/empirical/CI_ATE_dgp0_tuning#.eps (*Figure 1 in Section 6*)
- images/empirical/CI_ATT_dgp0_tuning#.eps (*Figure 1 in Section 6*)
- output/empirical/CI/CI_dgp0_tuning#_n445_sample1.mat
- output/empirical/CI/CIatt_dgp0_tuning#_n445_sample1.mat
- output/empirical/CI/CI_dgp0_tuning#_n390_sample2.mat
- output/empirical/CI/CIatt_dgp0_tuning#_n390_sample2.mat

2. `main_MC_coverageProb.m`

Description

- Coverage probability calculation for each fixed c.
- It first calls `getAteTrue.m` to compute the true ATE bounds and then calls the function `computeCI.m` to estimate confidence sets on ATE bounds. Results are then passed to `computeLargestDeviation.m` and `plotCoverageProb.m`

Inputs (customizable in the script `setChoiceMC.m` .)

- dgpChoiceVec: dgp choice index
- tuningChoiceVec: tuning parameter index
- alphaVec: significance level
- sampleSize

- nSimulations: number of replications

Main output files

- tables/CP_dgp#_tuning#_n#_alpha#.xls
- images/CI_dgp#_tuning#_n#_sim#_alpha#.eps
- images/CP_dgp#_tuning#_n#_sim#_alpha#.eps (*Figure S5 in Appendix H.2*)
- The largest deviation of coverage probabilities reported in Appendix H.2

3. `main_MC_misspec.m`

Description

- Performance evaluation of misspecification.
- It first calls `getAteTrue.m` to compute the true ATE bounds and then calls the function `computeATE.m` to estimate ATE bounds for three values of polynomial orders.
- Results saved are estimated IMSE (scripts involved: `computeStats.m` , `makeTableIMSE.m`) and a visual representation of the empirical distribution of the estimated bound functions (script involved: `plotCI.m`).

Inputs

- dgpChoiceVec: dgp choice index
- tuningChoice: tuning parameter index
- nSimulations: number of replications
- sampleSizeVec
- optimalR: number of nodes used for numerical integration (over the population distribution of W)

Main output files

- images/misspec_dgp#_tuning#_n#_poly#_sim#.eps (*Figure S4 in Appendix H.1*)
- output/fullMC/misspecification/imse_dgp#_tuning#.tex (*Table S1 in Appendix H.1*)

Stata program

The main file in the codeStata folder is `LaLondeReplication.do` . Run this to obtain Figure 2 and Tables 1-4 in section 6, and Table S2 in Appendix J.

Description

The program calls our companion Stata module `tesensitivity`. To install it, type `ssc install tesensitivity` from within Stata. Type `help tesensitivity` for syntax and instructions.

First it runs our module to estimate bounds on ATE and ATT under deviations from selection on observables. After estimating bounds on treatment effects, it implements the leave-out-variable-k propensity score analysis discussed in section 2 of the paper. This compares the propensity scores obtained when leaving out each covariate in the analysis compared to the propensity scores including all covariates. This provides a reference point for thinking about the level of c-dependence that might be caused by other omitted variables.

Finally, it runs the program leave_out_ate.ado. This program outputs the magnitude of the effect of omitting a single variable on ATE point estimates. It takes a list of outcome, treatment, and controls as an input, computes ATE using inverse probability weighting (IPW), and saves the IPW estimate as our main baseline. Next, for each control variable, it computes the IPW estimate without that control. This allows us to see the impact of adding that control on the ATE estimate. Variables which do not have a large marginal impact on the ATE estimate may be considered unimportant for calibrating our sensitivity analysis.