

Doubly Robust Difference-in-Differences Estimators: Readme File

Pedro H. C. Sant'Anna*

Jun B. Zhao†

Vanderbilt University

Vanderbilt University

May 5, 2020

1 Monte Carlo Simulation Replication

In order to replicate the Monte Carlo simulation study of the paper, one uses the files located in the folder **“MonteCarlo”**.

For the case of panel data, one refers to the folder **“panel”**. **“main.panel.r”** is the code to run all simulations, which calls the DGP function **“dgps.KS.R”**. **“main.panel.r”** uses **“sim.panel.R”**, where the functions in the DRDID R package are called to calculate all estimators:

“DRDID::drdid_panel” is used to calculate $\hat{\tau}^{dr,p}$;

“DRDID::drdid_imp_panel” is used to calculate $\hat{\tau}_{imp}^{dr,p}$;

“DRDID::reg_did_panel” is used to calculate $\hat{\tau}^{reg}$;

“DRDID::ipw_did_panel” is used to calculate $\hat{\tau}^{ipw,p}$;

“DRDID::std_ipw_did_panel” is used to calculate $\hat{\tau}_{std}^{ipw,p}$;

“DRDID::twfe_did_panel” is used to calculate $\hat{\tau}^{fe}$.

Besides, the corresponding simulated semiparametric efficiency bounds are calculated using **“main.panel.efficiency.R”**, which calls the DGP function **“dgps.KS.efficiency.R”**, and

*Department of Economics, Vanderbilt University. E-mail: pedro.h.santanna@vanderbilt.edu.

†Department of Economics, Vanderbilt University. E-mail: jun.zhao@vanderbilt.edu.

simulations are conducted through calling **“sim.panel.efficiency.R”**.

For the case of repeated cross-section data, one refers to the folder **“rc”**. **“main_rc.R”** is the code to run all simulations, which calls the DGP function **“dgps_KS_rc.R”**. **“main.panel.r”** uses **“sim_rc.R”**, where the functions in the DRDID R package are called:

“DRDID::drdid_rc1” is used to calculate $\hat{\tau}_1^{dr,rc}$;

“DRDID::drdid_imp_rc1” is used to calculate $\hat{\tau}_{1,imp}^{dr,rc}$;

“DRDID::drdid_rc” is used to calculate $\hat{\tau}_2^{dr,rc}$;

“DRDID::drdid_imp_rc” is used to calculate $\hat{\tau}_{2,imp}^{dr,rc}$;

“DRDID::reg_did_rc” is used to calculate $\hat{\tau}^{reg}$;

“DRDID::ipw_did_rc” is used to calculate $\hat{\tau}^{ipw,rc}$;

“DRDID::std_ipw_did_rc” is used to calculate $\hat{\tau}_{std}^{ipw,rc}$;

“DRDID::twfe_did_rc” is used to calculate $\hat{\tau}^{fe}$.

Besides, the corresponding simulated semiparametric efficiency bounds are calculated using **“main_rc_efficiency.R”**, which calls the DGP function **“dgps_KS_rc_efficiency.R”**, and simulations are conducted through calling **“sim_rc_efficiency.R”**.

2 Empirical Illustration Replication

In order to replicate the empirical illustration about the effect of job training on earnings, one uses the files located in the folder **“Application”**. The data is stored in the folder **“data”**, called **“nsw.dta”**.

Firstly, one needs to create the subsamples through calling **“subsamples.R”**, which generates three subsamples used in the paper: the LaLonde sample, the DW sample, and the early RA sample, named **“eval.lalonde.cps”**, **“eval.dw.cps”**, and **“eval.early.cps”** respectively. Secondly, one uses the function **“all.did.subsample”** in **“all.did.subsample.R”** to obtain the results for each of the three subsamples, which contains the three specifications considered in the paper. The function **“all.did.estimators”** in **“all.did.estimators.R”** is used to calculate all estimators. Finally, the estimated results are exported using the function **“out.table.lalonde”** in **“out.table.lalonde.R”**.