

# Replication Guide for “Double Robust Bayesian Inference on Average Treatment Effects”

Christoph Breunig, Ruixuan Liu and Zhengfei Yu

October 5, 2024

## Overview

The replication package contains R and MATLAB code. The R code draws the simulation samples, prepares the empirical data, estimates the propensity score, and implements all frequentist methods. The MATLAB code implements the Bayesian methods.

The folders `Table_1`, `Table_2` and `Table_Supp` contain the code used to replicate Table 1 (in Section 5.1), Table 2 (in Section 5.2), and the tables in the online supplementary materials (Appendix H), respectively. The folder `Matlab_func` contains auxiliary user-defined functions that are called when running the MATLAB code in other folders. Most of the user-defined functions are from the GPML toolbox Rassmussen and Williams [2006]. The folder `Data` contains two sets of raw data (to be detailed below).

## Data Availability and Provenance Statements

### Statement about Rights

I certify that the authors of the manuscript have legitimate access to and permission to use the data used in this manuscript.

### Summary of Availability

All data are publicly available.

## Details on Data Source

The folder `Data` contains the following two sets of raw data:

- For simulations: “psid\_generated.feather”, which is large pseudo-data pool generated by Athey et al. [2024] and can be downloaded from: <https://github.com/evanmunro/dswgan-paper/tree/main/data/generated#readme> (in the Google Drive folder).
- For empirical application: “nswre74\_treated.txt” (185 observations), “nswre74\_control.txt” (260 observations) and “psid\_controls.txt” (2490 observations). These data sets were initially used by LaLonde [1986] and Dehejia and Wahba

[1999], and can be downloaded from Dehejia’s website:  
<http://users.nber.org/~rdehejia/nswdata2.html>.

## Computational Requirements

### Software Requirements

We use R version 4.4.1 (2024-06-14) and MATLAB version 9.12.0.1884302 (R2022a).

For the R code, the following packages were used for implementation: `glmnet` (version 4.1.8), `Matching` (version 4.10.14), `drtmle` (version 1.1.2), `SuperLearner` (version 2.0.29), and `DoubleML` (version 1.0.1).

### Memory and Runtime Requirements

The code was run on a desktop machine (Intel(R) Xeon(R) CPU E5-2620 v4, 32GB RAM) with 64-bit Windows system (x86\_64-w64-mingw32).

Reproducing Table 1 takes about one day. The replicator can run Step 3 (MATLAB code) and Step 4 (R code) in Folder **Table\_1** simultaneously, see Instruction section for details. Reproducing Table 2 takes three minutes. Tables A1 and A2 in online appendix H takes about three days and one day to reproduce, respectively.

### Controlled Randomness

The random seed is set when drawing simulation samples from the pseudo-data pool generated by Athey et al. [2024]. For Bayesian methods, the random seed is set when drawing posteriors. For frequentist methods, random seeds are set when using the R package `drtmle` to compute the DR TMLE estimator (Benkeser et al. [2017]) and `DoubleML` to compute the DML estimator (Chernozhukov et al. [2017]).

## Instructions for Replicators

In the following, we provide step-by-step instructions for replicating each table in the paper. The output is provided separately for each trimming threshold  $t \in \{0.1, 0.05, 0.01\}$ .

### Folder **Table\_1**

The simulation results in Table 1 in Section 5.1 can be replicated as follows:

1. Draw simulation samples: `Data_draw.R`  
Use the pseudo-data “psid\_generated.feather” in folder **Data**.  
Outputs are saved as “Simu\_data\_psid\_n2675” in R and MATLAB data formats in folder **Data**.
2. Estimate the propensity score using the data drawn in Step 1: `PS_est.R`  
Outputs are saved as “PS\_est\_psid\_n2675” in R and MATLAB data formats in folder **Data**.

(This step also computes the optimal trimming threshold as stated in Footnote 5.)

3. Bayesian methods: `BDR_ATE.m`  
Replicates the results for Bayes, PA Bayes and DR Bayes in Table 1 using the simulation samples in Step 1 and PS estimates in Step 2.  
Results are saved as `Table_1.Bayes.txt`.
4. Frequentist methods: `FRE_ATE.R`  
Replicates the rows of methods Match, Match BC, DR TMLE and DML in Table 1 using the simulation samples in Step 1 and PS estimates in Step 2.  
Results are saved in `Table_1.Freq.txt`.

## Folder Table\_2

The empirical results presented in Table 2 in Section 5.2 can be replicated as follows:

1. Load data and estimate propensity score: `JobTrain_Data_PS.R`  
Load the txt data files in the folder `Data`.  
Outputs are saved as “`Real_ps_est.rda`” and “`Real_ps_est.mat`” in folder `Data`.  
(This step also computes the optimal trimming threshold  $t$  stated in Footnote 7.)
2. Bayesian methods: `Empirical_JobTrain.m`  
Replicates the rows of methods Bayes, PA Bayes and DR Bayes in Table 2 using the output of Step 1.  
Results are saved in `Table_2.Bayes.txt`.
3. Frequentist methods: `FRE_Empirical_JobTrain.R`  
Replicates the results for Match, Match BC, DR TMLE and DML in Table 2 using the output of Step 1.  
Results are saved in `Table_2.Freq.txt`.
4. The in-text numbers in the first paragraph of Section 5.2: (Point estimate and 95% confidence interval of the experimental benchmark): `JobTrain_RCT.R`.  
Results are saved in `Table_2.Freq.txt`.

## Folder Table\_Supp

The tables in the Online Supplement Appendix H can be replicated as follows:

### Table A1 in Online Supplement Appendix H

`BDR_ATE_Sigma.m`  
Replicates the results for varying  $c_\sigma$  in Table A1 using the outputs from Steps 1 and 2 of Table 1.  
The results are saved in `Table_A1.txt`.  
Note that the rows with  $c_\sigma = 1$  in Table A1 restate the rows of PA and DR Bayes in Table 1 for comparison purposes.

### Table A2 in Online Supplement Appendix H

1. Estimate the propensity score: PS\_est\_SS.R  
The outputs are saved as “PS\_est\_SS\_psid\_n2675.mat” in the folder Data.
2. Bayesian methods: BDR\_ATE\_SS.m  
Replicates the “Sample-split” section of Table A2 using the output from Step 1.  
Results are saved in Table\_A2.txt, which replicates the “Sample-split” section of Table A2. The “Full sample” section restates the rows of PA and DR Bayes in Table 1 for comparison purposes.

## References

- S. Athey, G. W. Imbens, J. Metzger, and E. Munro. Using wasserstein generative adversarial networks for the design of monte carlo simulations. *Journal of Econometrics*, 240(2):105076, 2024.
- D. Benkeser, M. Carone, M. V. D. Laan, and P. Gilbert. Doubly robust non-parametric inference on the average treatment effect. *Biometrika*, 104(4):863–880, 2017.
- V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, and W. Newey. Double/debiased/neyman machine learning of treatment effects. *American Economic Review*, 107(5):261–265, 2017.
- R. H. Dehejia and S. Wahba. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association*, 94(448):1053–1062, 1999.
- R. J. LaLonde. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, 76(4):604–620, 1986.
- C. Rasmussen and C. Williams. *Gaussian processes for machine learning*. MIT, 2006.