

Replication Files for “Minimum Resource Threshold Policy Under Partial Interference”

This supplementary file contains replication codes for “Minimum Resource Threshold Policy Under Partial Interference.”

1 Software and Packages

- Software: R version 4.2.1
- Packages: caret (version 6.0.93); dplyr (version 1.1.2); earth (version 5.3.1); GADMTTools (version 3.9.1); gam (version 1.22); gbm (version 2.1.8.1); ggplot2 (version 3.4.2); ggpubr (version 0.5.0); glmnet (version 4.1.6); grid (version 4.2.1); gridExtra (version 2.3); gtools (version 3.9.4); kernlab (version 0.9.32); KernSmooth (version 2.23.20); lattice (version 0.20.45); maps (version 3.4.1); MASS (version 7.3.58.2); nnet (version 7.3.18); np (version 0.60.16); polyspline (version 1.1.22); ranger (version 0.14.1); RcolorBrewer (version 1.1.3); rgdal (version 1.6.4); SuperLearner (version 2.0.28); xgboost (version 1.7.3.1); readstata13 (version 0.10.0)

2 Common Code Files

- MySL.R contains functions used for implementing superlearner algorithm and estimating the nuisance functions.
- OptTrt_Sim_Function.R contains functions used for the MRTR using the direct and indirect approaches.

3 Data Folder

Data folder contains replication files for the data analysis in Section 5 of the paper.

3.1 Download the DHS data

- The dataset can be downloaded by following the steps below:
 - Request dataset access by following the instructions at <https://dhsprogram.com/data/Access-Instructions.cfm>.

- Download SNKR70FL.ZIP from https://dhsprogram.com/data/dataset/Senegal_Continuous-DHS_2014.cfm?flag=1.
 - Download SNKR7HFL.ZIP from https://dhsprogram.com/data/dataset/Senegal_Continuous-DHS_2015.cfm?flag=1.
 - Download SNKR7IFL.ZIP from https://dhsprogram.com/data/dataset/Senegal_Continuous-DHS_2016.cfm?flag=1.
 - Download SNKR7ZFL.ZIP from https://dhsprogram.com/data/dataset/Senegal_Continuous-DHS_2017.cfm?flag=1.
 - Download SNKR81FL.ZIP from https://dhsprogram.com/data/dataset/Senegal_Continuous-DHS_2018.cfm?flag=1.
 - Unzip the above zip files, and obtain SNKR70FL.DTA, SNKR7HFL.DTA, SNKR7IFL.DTA, SNKR7ZFL.DTA, SNKR81FL.DTA.
- Place the datasets in the following paths.
 - Data/DHS/2014/SNKR70FL.DTA
 - Data/DHS/2015/SNKR7HFL.DTA
 - Data/DHS/2016/SNKR7IFL.DTA
 - Data/DHS/2017/SNKR7ZFL.DTA
 - Data/DHS/2018/SNKR81FL.DTA
 - Download “SNGE81FL.ZIP” from https://dhsprogram.com/data/dataset/Senegal_Continuous-DHS_2018.cfm?flag=1.
 - Unzip the zip file and place the 10 files below in “Data/GPS_2018” folder: DHS_README.txt, GPS_Displacement_README.txt, SNGE81FL.cpg, SNGE81FL.dbf, SNGE81FL.prj, SNGE81FL.sbn, SNGE81FL.sbx, SNGE81FL.shp, SNGE81FL.shp.xml, SNGE81FL.shx.

3.2 Code

- Step1_Cleaning.R
 - The DHS datasets are cleaned.
 - The training and test datasets are saved as “Sen_Training_1417.csv” and “Sen_Test_18.csv”, respectively.
 - The test set including the identifier is saved as “TestSetID.csv”.
- Step2_MakeSplittingID.R
 - Cross-fitting sets are saved in SAM_IND.csv and SS_IND.csv
- Step3_Direct_NF_Estimation.R

- The outcome regressions and propensity scores are estimated.
- The results are saved as “NF_SS###_ES###.csv” and “PS_SS###_ES###.csv” files in “NF” folder.
- This step will take a lot of time, so it is recommended to use parallel computing.
- Step4_Direct_NF_Merge.R
 - Files obtained in Step 3 are merged.
 - The results are saved as “NF_SS###.csv” and “PS_SS###.csv” files in “NF” folder.
- Step5_Direct_OV_InitialPoints.R and Step5_Direct_SO_InitialPoints.R
 - The initial values for the DC algorithm are evaluated.
 - The results are saved as “Starting_B#####.csv” files in “Overall/Starting” and “Spillover-all/Starting” folders.
- Step6_Direct_OV_InitialPoints_Adjustment.R and Step6_Direct_SO_InitialPoints_Adjustment.R
 - Adjust the initial points obtained in the Step 5 R files.
 - The results are saved as “NF_Train_Adj_B#####.csv” files in “Overall/NF_New_Adj” and “Spilloverall/NF_New_Adj” folders.
- Step7_Direct_OV_SVMPParameter.R and Step7_Direct_SO_SVMPParameter.R
 - The empirical risks under each parameter are evaluated.
 - The results are saved as “CV_OV_S#####.csv” files in “Overall/CV” and “Spilloverall/CV” folders.
 - This step will take a lot of time, so it is recommended to use parallel computing.
- Step8_Direct_OV_TrainingsetRule.R and Step8_Direct_SO_TrainingsetRule.R
 - MRTR estimates obtained from the direct approach for the training sets are obtained where parameters are chosen from CV.
 - The results are saved as “RULE_S#####.csv” files in “Overall/RULE” and “Spilloverall/RULE” folders.
 - This step will take a lot of time, so it is recommended to use parallel computing.
- Step9_Direct_OV_TestsetRule.R and Step9_Direct_SO_TestsetRule.R
 - MRTR estimates obtained from the direct approach for the test sets are obtained where parameters are chosen from CV.
 - The results are saved as “RULE_S#####-Test.csv” files in “Overall/RULE” and “Spilloverall/RULE” folders.

- Step10_Indirect_Estimation.R
 - MRTR estimates obtained from the indirect approach for the test sets are obtained where parameters are chosen from CV.
 - The results are saved as “RULE_LR_Test.csv” files in “Overall/NF_LR” and “Spilloverall/NF_LR” folders.
- Step11_Summary_GeoRegion.R
 - The simulation results are merged with GPS file “Data/GPS_2018/SNGE81FL.shp”
 - The summary is saved as “Overall/RULE_Merged.csv” and “Spilloverall/RULE_Merged.csv”.
- Step12_CheckAssumptions.R
 - The assumptions are assessed as in Section 5.2.
- Step13_Summary.R
 - The simulation results are summarized and the plots are drawn.

3.3 Folder

- Data: this folder contains the cleaned datasets used in the analysis.
- Data.Original: this folder contains 2014-2018 Senegal DHS datasets and the GPS information of 2018 Senegal DHS dataset.
- NF: this folder contains the estimated outcome regression and propensity score by aggregating the files in NF_Raw folder.
- NF_LR: this folder contains MRTR estimates obtained from the indirect approach.
- NF_Raw: this folder contains the estimated outcome regression and propensity score obtained from undersampling.
- Overall: this folder contains the results related to the overall MRTR estimation.
- plot: this folder contains the graphical summaries of the data analysis and the assessment of the assumptions.
- Spilloverall: this folder contains the results related to the spilloverall MRTR estimation.
- SplittingID: this folder contains the indices of cross-fitting procedure.

4 Simulation Folder

Simulation folder contains replication files for the simulation in Section 4 of the paper.

4.1 Code

- Step1_DGP.R
 - The datasets used for the simulation are generated.
- Step2_Direct_NF_Estimation.R
 - The outcome regressions and propensity scores are estimated.
 - The results are saved as “NF_TT###_SS###_ES###.csv” files in “NF_Raw” folder.
 - This step will take a lot of time, so it is recommended to use parallel computing.
- Step3_Direct_NF_Merge.R
 - Files obtained in Step 3 are merged.
 - The results are saved as “NF_TT###_SS###.csv” files in “NF” folder.
- Step4_Direct_InitialPoints.R
 - The initial values for the DC algorithm are evaluated.
 - The results are saved as “Starting_TT###_SS###.csv” files in “Starting” folder.
- Step5_Direct_InitialPoints_Adjustment.R
 - Adjust the initial points obtained in the Step 4 R file.
 - The results are saved as “NF_Train_Adj_TT###_SS###.csv” files in “NF” folder.
- Step6_Direct_SVMPParameter.R
 - The empirical risks under each parameter are evaluated.
 - The results are saved as “CV_TT.S###.csv” files in “CV” folder.
 - This step will take a lot of time, so it is recommended to use parallel computing.
- Step7_Direct_TrainingsetRule.R
 - MRTR estimates obtained from the direct approach for the training sets are obtained where parameters are chosen from CV.
 - The results are saved as “RULE_TT###_S###.csv” files in “RULE” folder.
 - This step will take a lot of time, so it is recommended to use parallel computing.
- Step8_Direct_TestsetRule.R
 - MRTR estimates obtained from the direct approach for the test sets are obtained where parameters are chosen from CV.
 - The results are saved as “RULE_TT###_S###_Test.csv” files in “RULE” folder.

- Step9_Indirect_Estimation.R
 - MRTR estimates obtained from the indirect approach for the test sets are obtained where parameters are chosen from CV.
 - The results are saved as “LR_OR_Test_TT###.csv” and “RULE_LR_Test_TT###.csv” files in “Indirect” folder.
- Step10_Summary.R
 - The simulation results are summarized.

4.2 Folder

- CV: this folder contains cross-validation results.
- Data: this folder contains the datasets generated from the simulation scenario.
- Indirect: this folder contains the result of the MRTR estimation using the indirect approaches.
- NF: this folder contains the estimated outcome regression and propensity score by aggregating the files in NF_Raw folder.
- NF_Raw: this folder contains the estimated outcome regression and propensity score obtained from undersampling.
- RULE: this folder contains the result of the MRTR estimation using the direct approaches.
- plot: plot folder contains the graphical summaries of the data analysis and the assessment of the assumptions.
- Starting and Starting_Raw: these folders contain the initial points for the empirical risk minimization.