README for

"Selection with Variation in Diagnostic Skill: Evidence from Radiologists"

1 Organization

The replication folder contains files necessary to reproduce all tables and figures used within the paper. The sub-folders are structured as follows:

- `source/`: code files used to clean and analyze data, estimate models, and create tables and figures.
  - `setup/`: code files used to install required packages.
  - `lib/`: R functions and Stata globals used in this replication package.
  - `data/`: code files used to clean and analyze data.
  - `estimate/`: code files used to estimate models.
  - `analysis`: code files used to create tables and figures.
  - `bootstrap/`: code files used to bootstrap standard errors for various estimates.
  - `figures/`: manually created figures.
  - `tables/`: manually created tables, code files used to format table inputs, and LyX templates.
- `build/`: intermediate and final output files created by code files in `source`.
  - `data/`: output files created by `source/data/`.
  - `estimate/`: output files created by `source/estimate/`.
  - `analysis`: output files created by `source/analysis/`.
  - `bootstrap/`: output files created by `source/bootstrap/`.
  - `tables/`: output files created by `source/tables/`.
- `figures_tables/`: final figures and tables renamed to reflect their locations in the paper.

  - `make.py`: code file used to build tables and rename figures.

- `replication_files.xlsx`: A complete list of both code and output files in this package necessary to reproduce all tables and figures, together with a short description for each file and whether it is included in this replication package. For each output file under `build/`, the last column indicates the code that produces it.

2 Data

This paper studies radiologists' pneumonia diagnoses using data from the Veterans Health Administration (VHA). All underlying data used in this paper include both patient and physician identifying information and are therefore not available for public release. The data include information about patients, such as demographics, vital status, medical history, and diagnoses, and information about radiologists, including demographics and diagnosis history. Access to the data requires employment by the VA. Interested readers may reach out to the authors for guidance on how to obtain access to the data. For readers who have access to VA data, the authors will assist with any reasonable replication attempts for two years following publication.

More specific information about the data needed to run all code files but may not be included directly in this replication package is provided in `replication_files.xlsx`. We include in this spreadsheet the date that any source data were created for us by the VA. We include processed data not containing identifying information about patients and physicians. Because all of the data included in the replication package are processed, replicators who have access to the underlying data do not need any data that we include in the package.

Statement about Rights

We certify that we have legitimate access to and permission to use the data used in this manuscript. We also certify that we have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission is documented in the LICENSE.txt file.

3 Execution

We execute all scripts using hardware and software in the VA Informatics and Computing Infrastructure (VINCI) environment. All replicators who have obtained access to the VA data will execute scripts in the same environment.

The execution of the code can be divided into several code blocks. Different code blocks should be run in the order that they appear on the list below. If not otherwise specified, within each code block, code files can be run in any order. For brevity, we omit the prefix `source/` from all code files below.

1. `setup/`: Install required packages.
    - `setup.do`
    - `setup.r`
2. `data/`: Construct the main dataset.
    - `iteration_SQLC.r` (will call `Rad_Sample_All.sql`)
3. `data/`: Construct auxiliary datasets.
    - `Rad_InpatientAdmission.sql`
    - `Rad_ExportLocationSID.sql`
    - `dual_enrollees_ED.sas`
    - `dual_enrollees_IP.sas`
    - `dual_enrollees_OP.sas`
    - `SQLView_Outpat.sql`
    - `SQLView_ED.sql`
    - `SQLView_Inpat.sql`
    - `Rad_Elixhauser.sql`
    - `Rad_Staff_Characteristics.sql`
4. `data/Rad_Prior_Util.sql`: Construct prior utilization.
5. `data/Rad_Sample_All.do`: Pull data from SQL servers and generate variables.
6. `data/Rad_GenerateData.do`: Combine all datasets and create the benchmark dataset.
7. `data/Rad_SelectStation.do`: Select stations with balance on age.
8. `data/`: Perform reduced-form analysis.
    - `Rad_EstimateSlope.do`
    - `Rad_EstimateSlopeByStation.do`
    - `Rad_EstimateSlopeOtherOutcomes.do`

- Rad_Balance.do
- Rad_Residualize.do
- Rad_Monotonicity.do
- Rad_RI.do
- Rad_Stability.do
- Rad_RF_Predictions.do
- Rad_RF_Predictions.r (should be run after Rad_RF_Predictions.do)

9. data/: Perform more reduced-form analysis.
    - Rad_GenerateMoments.do
    - Rad_Robustness.do
    - Rad_BalanceTable.do
10. estimation/: Estimate the baseline and alternative model specifications.
    - EstimateMainLambda.r
    - EstimateRobustnessAdmitted.r
    - EstimateRobustnessNoAdjustment.r
    - EstimateRobustnessNox.r
    - EstimateRobustnessSubset.r
    - EstimateRobustnessVAuser.r
    - EstimateMainFixLambda.r (should be run after EstimateMainLambda.r)
    - EstimateAge.r
    - EstimateMainVaryingZeta.r
11. analysis/: Create figures and tables.
    - Rad_PlotStatistics.do
    - Rad_PlotFirstStage.do
    - Rad_PlotRandomAssignment.do
    - Rad_PlotBalance.do
    - Rad_PlotMoments.do
    - Rad_PlotSlope.do

- Rad_PlotStability.do
- PlotEstimation.r
- PlotROC_bw.r
- PlotROC_robustness.r
- PlotVariations.r
- InvestigateInaccurateBelief.r
- PlotAUC.r
- PrintRates.r
- PlotWelfare.r
- SummarizeRobustness.r
- PrepareHeterogeneity.r
- Rad_PlotHeterogeneity.do (should be run after PrepareHeterogeneity.r)

12. analysis/: Perform a Monte Carlo exercise for controlling skill (should be run sequentially).
    - ControlSkill.r
    - Rad_PlotControlSkill.do
13. analysis/: Perform a Monte Carlo exercise for linear risk-adjustment (should be run sequentially).
    - SimulateRiskAdjustment.r
    - Rad_GenerateMomentsRiskAdjustment.do
    - EstimateRiskAdjustment.r
    - Rad_PlotRiskAdjustment.do
14. bootstrap/Rad_GenerateCounts.do: Prepare for drawing bootstrapped samples.
15. bootstrap/Rad_DrawBootstrapSample.r: Draw bootstrapped samples.
16. bootstrap/: Perform risk adjustment in bootstrapped samples.
    - GenerateMomentsMain.do
    - GenerateMomentsAdmitted.do
    - GenerateMomentsNoAdjustment.do
    - GenerateMomentsNox.do

- GenerateMomentsSubset.do
- GenerateMomentsVAuser.do

17. bootstrap/: Estimate models in bootstrapped samples.
    - EstimateBootstrapMainLambda.r
    - EstimateBootstrapRobustnessAdmitted.r
    - EstimateBootstrapRobustnessNoAdjustment.r
    - EstimateBootstrapRobustnessNox.r
    - EstimateBootstrapRobustnessSubset.r
    - EstimateBootstrapRobustnessVAuser.r
    - EstimateBootstrapMainFixLambda.r (should be run after EstimateBootstrapMainLambda.r)
18. bootstrap/: Bootstrap standard errors for various counterfactual estimates.
    - BootstrapPrimitives.r
    - BootstrapVarianceDecomposition.r
    - BootstrapStandardErrorsWelfare.r
    - Rad_Prediction_station.do
19. tables/Rad_FormatTable.do: Format table inputs.