Replication files for 'Finite-Sample Optimal Estimation and Inference on Average Treatment Effects Under Unconfoundedness'

Timothy B. Armstrong and Michal Kolesár

November 10, 2020

Instructions for replicating empirical results

1. Install the package `ATEHonest`, included in this replication. The package can also be installed from GitHub using

```
if (!requireNamespace("remotes")) {
  install.packages("remotes")
}
remotes::install_github("kolesarm/ATEHonest")
```

All functions in the package are documented, and the package also contains a vignette illustrating their use. The package also contains the NSW dataset used in the paper.

To re-create the plots, you also need to install the `ggthemr` package, using

```
remotes::install_github("cttobin/ggthemr")
```

The remaining package dependencies can be installed from CRAN.

2. To replicate the empirical results, run

```
rmarkdown::render("nsw_application.Rmd")
```

This generates:

- `nsw_application.pdf`: pdf file based on the R markdown file `nsw_application.Rmd` that contains all figures, tables, and calculations mentioned in the text of the paper. To keep the intermediate files used for generating the `.pdf`, use the option `clean=FALSE`
- object `t2`: Table 2
- `fig/nsw_opt.tex` Figure 1 in tikz format
- `fig/nsw_Lindw.tex` Plot of Lindeberg weights for the matching and optimal estimator (not reported in the paper)
- auxilliary files used to generate the `.pdf` file: `nsw_application.tex` source file, and the file `nsw_application-tikzDictionary`.

The `output/` directory contains the solution path for the main specification under `nsw_1path_3000.rda`. To compute all results from scratch, remove this file from the `output/` directory. The script then re-create them; this takes about 26 mins. Most of this time is spent on computing CIs in Figure 1 for very small values of C. To speed up the computation to about 4 mins, and only compute CIs for $C \geq 0.2$, set `ns <- 1100` on line 2 and `Cs <- c(seq(0.2, 4, by=0.1))` on line 3 of `nsw_1nonexper.R`