Factorial designs, model selection, and (incorrect) inference in randomized experiments

README File

Karthik Muralidharan Mauricio Romero Kaspar Wuthrich

16 enero, 2023

This file contains general instructions on what these codes do, and their sequence.

First, you need to run the STATA do-files. There is a master file in STATA that calls on all the other do-files. These do-files mostly deal with re-analyzing the data from previous papers. The STATA version we used to run the analysis was 16.1, but the do-files that re-analyze previous version may "set" different versions, in accordance with the replication do-files. You may need to install necessary STATA commands required by the replication files of different papers.

After the STATA do-files, you should run the R-code. The code in `RCode/PreviousPapers` takes the results from the re-analysis of previous papers and process it. You should run the three source files in this folder after the STATA do-files. The R version we used to run the analysis was 4.2.2.

Thee code in `RCode/Simulations` is the one that estimates the size and power for the different econometric approaches to deal with factorial designs. There is a master file that will call on all the other source files. This master file will also install all the necessary packages. We would like to thank Tim Armstrong, Michal Kolesar, and Soonwoo Kwon who graciously answered questions about the econometric method they developed (see https://arxiv.org/abs/2012.14823) and how to implement them. They provided code against which we checked our own implementation. Likewise, Adam McCloskey answered questions about the econometric method he developed (see https://www.sciencedirect.com/science/article/abs/pii/S0304407617300556 and https://www.tandfonline.com/doi/abs/10.1080/07350015.2019.1592754) throughout the years and provided code to us so that we could implement the consistent model selection, which we adapted to our setting. Finally, Graham Elliott was very kind in answering questions about the econometric methods he and his coauthors developed (see https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA10535). We based our code off the replication package of their paper (see https://www.econometricsociety.org/publications/econometrica/2015/03/01/nearly-optimal-tests-when-nuisance-parameter-present-under-null).

Finally, the code `RCode/TypeM_TypeS.r` estimates the type-M error figures in the paper. This code can be run independently of all other codes.

All the analyses were run using a Windows 10 machine (16GB of RAM, processor Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz 1.99 GHz, and a 1TB SSD). The run time is roughly 5 minutes for the Stata code. The R code to analyze the results from the replications takes less than 1 minute to run. The R code to do the simulations takes about 24 hours to run. The R code to do the type-M figures takes less than 1 minute.

1 Folders structure

The raw data in this project is in the `Crosscut Experiments` folder. This has summary files on the papers published in the `Top 5` journals in economics from 2007 to 2017

The folder `StataCode` has the STATA code to replicate the results of the paper. The `StataCode/00_Master.do` calls all the other do-files. You only need to change the global `folder`. In the `Crosscut Experiments/Papers`

folder, there is a folder for each one of the paper we re-analyze by changing whether they include the interaction terms or not. These folders have the replication data from the paper, and a do-file that re-analyzes the data.

The folder `RCode` has the R source files to replicate the results of the paper. As mentioned above, you should run the STATA code before running the RCode. The are three different part to the R code. First, the code in `RCode\PreviousPapers` which processes the results from the re-analysis of previous papers. You need to change the `setwd` in these files. Second, the code in `RCode/Simulations` is the one that estimates the size and power for the different econometric approaches to deal with factorial designs. There is a master file that will call on all the other source files. You need to change the `setwd` to the replication folder in your computer. Finally, the code `RCode/TypeM_TypeS.r` estimates the type-M error figures in the paper. You also need to change the `setwd`.

All the outputs (e.g., `.tex` files and figures) go to the folder `LaTeX`. There is a sub folder for the simulations (`LaTeX/plots_sims`), one for the replications (`LaTeX/replicate_summary`), and one for figures and tables in general (`LaTeX/figures`and `LaTeX/tables`).

2 Recreating the final tables and figures

For convenience we have placed the source .tex files so that anyone can re-compile the paper with figures/tables. See `3 Replication package/LaTeX/FiguresAndTables.TeX`.

However, here are the output files that correspond to each figure in the paper and the appendix.

- Table 1: Does not have any statistical analysis in it
- Table 2: Does not have any statistical analysis in it
- Table 3, Panel A corresponds to `replicate_summary/conf_10.tex`
- Table 3, Panel B corresponds to `replicate_summary/conf_5.tex`
- Table 3, Panel C corresponds to `replicate_summary/conf_1.tex`
- Figure 1, Panel A corresponds to `plots_sims/Figure1_type2_2_1000_slides.pdf`
- Figure 1, Panel B corresponds to `plots_sims/Figure1_Bias_1000.pdf`
- Figure 1, Panel C corresponds to `plots_sims/Figure1_type1_1000.pdf`
- Figure 2, corresponds to `plots_sims/Figure2_type1_1000_extended.pdf`
- Figure 3, corresponds to `replicate_summary/histogram_int_trimmed01_fancy.pdf`
- Figure 4, Panel A corresponds to `replicate_summary/ScatterLongShort.pdf`
- Figure 4, Panel B corresponds to `replicate_summary/ScatterLongShort_trimmed.pdf`
- Figure 5, Panel A corresponds to `plots_sims/Figure4_type1_1000.pdf`
- Figure 5, Panel B corresponds to `plots_sims/Figure4_type2_1000_2_oneside.pdf`
- Figure 6, Panel A corresponds to `plots_sims/Figure_AK_type1_1000_compare.pdf`
- Figure 6, Panel B corresponds to `plots_sims/Figure_AK_type2_2_1000_compare.pdf`
- Figure 7, Panel A corresponds to `plots_sims/Figure_Compare_type1_1000.pdf`
- Figure 7, Panel B corresponds to `plots_sims/Figure_Compare_type2_2_1000.pdf`

In the Appendix, the figures and tables correspond to the following files.

- Table A1: Does not have any statistical analysis in it
- Table A2, Panel A corresponds to `replicate_summary/conf_10Top10.tex`

- Table A2, Panel B corresponds to `replicate_summary/conf_5Top10.tex`
- Table A2, Panel C corresponds to `replicate_summary/conf_1Top10.tex`
- Table A3, Panel A corresponds to `replicate_summary/conf_10_policy.tex`
- Table A3, Panel B corresponds to `replicate_summary/conf_5_policy.tex`
- Table A3, Panel C corresponds to `replicate_summary/conf_1_policy.tex`
- Table A4, Panel A corresponds to `replicate_summary/conf_10_Included.tex`
- Table A4, Panel B corresponds to `replicate_summary/conf_5_Included.tex`
- Table A4, Panel C corresponds to `replicate_summary/conf_1_Included.tex`
- Table A5: Does not have any statistical analysis in it
- Table A6: corresponds to `tables/CrossTab.tex`
- Figure A1, corresponds to `replicate_summary/Histogram_tvalue.pdf`
- Figure A2, Panel A corresponds to `replicate_summary/ScatterLongShortTop10.pdf`
- Figure A2, Panel B corresponds to `replicate_summary/histogram_int_trimmed01_fancy_Top10.pdf`
- Figure A3, corresponds to `replicate_summary/Histogram_tvalueTop10.pdf`
- Figure A4, Panel A corresponds to `replicate_summary/ScatterLongShort_policy.pdf`
- Figure A4, Panel B corresponds to `replicate_summary/histogram_int_trimmed01_fancy_policy.pdf`
- Figure A5, corresponds to `replicate_summary/Histogram_tvalue_policy.pdf`
- Figure A6, Panel A corresponds to `replicate_summary/ScatterLongShort_Included.pdf`
- Figure A6, Panel B corresponds to `replicate_summary/Histogram_Int_Trimmed01_Included_fancy.pdf`
- Figure A7, corresponds to `replicate_summary/Histogram_tvalue_Included.pdf`
- Figure A8, corresponds to `figures/Relative_SampleSize_05_single.pdf`
- Figure A9, corresponds to `figures/typeM_05_single.pdf`
- Figure A10, Panel A corresponds to `plots_sims/Figure_IMS_type1_1000_twosided.pdf`
- Figure A10, Panel B corresponds to `plots_sims/Figure_IMS_type2_2_1000_twosided.pdf`
- Figure A11, Panel A corresponds to `plots_sims/Figure_ConsistentMS_type1_1000.pdf`
- Figure A11, Panel B corresponds to `plots_sims/Figure_ConsistentMS_type2_2_1000.pdf`
- Figure A12, Panel A corresponds to `plots_sims/Figure1_PowerTrace2_100.pdf`
- Figure A12, Panel B corresponds to `plots_sims/Figure1_PowerTrace2_500.pdf`
- Figure A12, Panel C corresponds to `plots_sims/Figure1_PowerTrace2_1000.pdf`
- Figure A12, Panel D corresponds to `plots_sims/Figure1_PowerTrace2_10000.pdf`
- Figure A13, Panel A corresponds to `plots_sims/Figure_MS_PowerTrace2_100.pdf`
- Figure A13, Panel B corresponds to `plots_sims/Figure_MS_PowerTrace2_500.pdf`
- Figure A13, Panel C corresponds to `plots_sims/Figure_MS_PowerTrace2_1000.pdf`
- Figure A13, Panel D corresponds to `plots_sims/Figure_MS_PowerTrace2_10000.pdf`
- Figure A14, Panel A corresponds to `plots_sims/Figure4_PowerTrace2_100.pdf`
- Figure A14, Panel B corresponds to `plots_sims/Figure4_PowerTrace2_500.pdf`

- Figure A14, Panel C corresponds to `plots_sims/Figure4_PowerTrace2_1000.pdf`
- Figure A14, Panel D corresponds to `plots_sims/Figure4_PowerTrace2_10000.pdf`
- Figure A15, Panel A corresponds to `plots_sims/Figure_AK_PowerTrace2_100.pdf`
- Figure A15, Panel B corresponds to `plots_sims/Figure_AK_PowerTrace2_500.pdf`
- Figure A15, Panel C corresponds to `plots_sims/Figure_AK_PowerTrace2_1000.pdf`
- Figure A15, Panel D corresponds to `plots_sims/Figure_AK_PowerTrace2_10000.pdf`
- Figure A16, Panel A corresponds to `plots_sims/Figure_IM_PowerTrace2_100.pdf`
- Figure A16, Panel B corresponds to `plots_sims/Figure_IM_PowerTrace2_500.pdf`
- Figure A16, Panel C corresponds to `plots_sims/Figure_IM_PowerTrace2_1000.pdf`
- Figure A16, Panel D corresponds to `plots_sims/Figure_IM_PowerTrace2_10000.pdf`
- Figure A17, Panel A corresponds to `plots_sims/Figure_NoInter_Optim_PowerTrace2_100.pdf`
- Figure A17, Panel B corresponds to `plots_sims/Figure_NoInter_Optim_PowerTrace2_500.pdf`
- Figure A17, Panel C corresponds to `plots_sims/Figure_NoInter_Optim_PowerTrace2_1000.pdf`
- Figure A17, Panel D corresponds to `plots_sims/Figure_NoInter_Optim_PowerTrace2_10000.pdf`
- Figure A18, Panel A corresponds to `plots_sims/Figure_Consistent_MS_PowerTrace2_100.pdf`
- Figure A18, Panel B corresponds to `plots_sims/Figure_Consistent_MS_PowerTrace2_500.pdf`
- Figure A18, Panel C corresponds to `plots_sims/Figure_Consistent_MS_PowerTrace2_1000.pdf`
- Figure A18, Panel D corresponds to `plots_sims/Figure_Consistent_MS_PowerTrace2_10000.pdf`
- Figure A19, Panel A corresponds to `plots_sims/Figure_Compare_type1_100.pdf`
- Figure A19, Panel B corresponds to `plots_sims/Figure_Compare_type2_2_100.pdf`
- Figure A19, Panel C corresponds to `plots_sims/Figure_Compare_type1_500.pdf`
- Figure A19, Panel D corresponds to `plots_sims/Figure_Compare_type2_2_500.pdf`
- Figure A19, Panel E corresponds to `plots_sims/Figure_Compare_type1_1000.pdf`
- Figure A19, Panel F corresponds to `plots_sims/Figure_Compare_type2_2_1000.pdf`
- Figure A19, Panel G corresponds to `plots_sims/Figure_Compare_type1_10000.pdf`
- Figure A19, Panel H corresponds to `plots_sims/Figure_Compare_type2_2_10000.pdf`