

READ ME FILE

Title: Star Wars at Central Banks

Authors: Adam Gorajek, Joel Bank, Andrew Staib, Benjamin Malin and Hamish Fitchett

Description

This 'read me' file contains general instructions on how to replicate the results, tables and graphs in RDP 2021-02 using the code and data released along with the main paper. 'final_paper_graph_data.xlsx' contains the data used to plot the figures in the main paper in an excel format.

If you make use of any of these files you should clearly attribute the authors in any derivative work.

General replication instructions

The replication programs are in the 'Programs' folder, in self-explanatory folder names. Descriptions of the individual programs in the next section. Together, these programs are adaptations of those provided by Simonsohn, Nelson and Simmons (2014) and Brodeur *et al* (2016).

To preserve relative file paths:

- execute all Stata programs through Stata project 'Star Wars at Central Banks.stpr'
- execute all R programs through R project 'Star Wars at Central Banks.Rproj'

To reproduce all analysis, work in alphabetical order of program folders and, within each folder, in numbered order of programs.

There are two package dependencies in 'C1_export_do'. To install the packages, run the following two commands in Stata:

- capture : ssc install kdens //For kernel densities with boundary adjustments
- capture : ssc install moremata //dependency for kdens package

There are two package dependencies in 'C2_1_estimation_non_param.R'. To install the packages, run the following two commands in R:

- install.packages("foreign")
- install.packages("isotone")

There are several package dependencies to run the p-curve analysis and create the p-curve graphs (see all programs with the prefix 'D'). To install the packages, run the following commands in R:

- install.packages("ggplot2")
- install.packages("ragg")
- install.packages("tidyverse") – we suggest installing the complete tidyverse to make things easy if you haven't already
- install.packages("haven")
- install.packages("vctrs")

There are several packages dependencies to construct the z-curve graphs that are displayed in the paper (see programs with the prefix 'E'):

- install.packages("ggplot2")
- install.packages("ragg")
- install.packages("tidyverse")

- `install.packages("haven")`
- `install.packages("poibin")`

We used STATA version 16.1 and R version 3.6.1.

The versions of each STATA package used are:

- `kdens: 2.0.2`

The versions of each R package used are:

- `isotone: 1.1-0`
- `foreign: 0.8-71`
- `ggplot2: 3.3.2`
- `ragg: 0.3.1`
- `tidyverse: 1.3.0`
- `haven: 2.2.0`
- `vctrs: 0.2.4`
- `poibin: 1.5`

Note: we are aware of an issue caused by using version 0.4.1 of the “ragg” package. The issue is that the png files of paper-formatted graphs do not render correctly. We are not aware of any other package dependency issues.

Program descriptions

A1_import_stats.do

Summary: Starts from disaggregated collected data and constructs a unique dataset.

Inputs: Sheets of various ‘DataCollection*.xlsx’ files stored in the ‘Data/Source/raw’ directory. Each file corresponds to a data collection from a different author or research assistant. The ‘DataCollectionArellano.xlsx’ files comes jointly from co-author Malin and research assistant Arellano.

Outputs: Data/Temp/stars_raw

A2_import_meta_data.do

Summary: Imports collected meta-data about paper characteristics etc.

Inputs: Sheets of various ‘DataCollection*.xlsx’ files stored in the ‘Data/Source/raw’ directory.

Outputs: Data/Temp/supp_data_article

A3_final.do

Summary: Starts from raw imported data and creates the final well-organised dataset.

Inputs: Data/Temp/stars_raw

Data/Temp/supp_data_article

Data/Source/inputs/brodeur_final_stars_supp

Outputs: Data/Final/final_stars_supp

Notes: Several parts of this do-file create subsample dummies (which will equal 1 for each subsample in all other do-files).

B1_collate_author_data.do

Summary: Starts from disaggregated collected data and constructs a dataset containing information about the authors of each central bank paper.

Inputs: Sheets of various 'DataCollection*.xlsx' files stored in the 'Data/Source/raw' directory.

Outputs: Data/Temp/author_x_article

B2_external_authors.do

Summary: Creates a dummy variable that shows if a given paper has at least one external author.

Inputs: Data/Temp/author_x_article

Outputs: Data/Temp/external_authors

B3_descriptive_statistics.do

Summary: Calculates descriptive statistics for the data.

Inputs: Data/Final/final_stars_supp
Data/Temp/external_authors

Outputs: None

Notes: The results from this script are used to populate Table 1 and inform the text in the main paper. But no physical outputs are produced.

C1_export.do

Summary: Generates the histogram and kernel density data for subgroups, later used to estimate selection functions. The simu file inputs are outputs of Brodeur *et al* (2016).

Inputs: Data/Final/final_stars_supp
Data/Source/inputs/simu_wdi
Data/Source/inputs/simu_psid
Data/Source/inputs/simu_vhlss
Data/Source/inputs/simu_qog

Outputs: Data/Temp/export_to_R.txt
Data/Temp/export_param

Notes: Uses STATA packages "kdens" and "moremata".

C2_1_estimation_non_param.R

Summary: Estimates the non-parametric selection function on various samples and returns residuals.

Inputs: Data/Temp/export_to_R.txt

Outputs: Data/Temp/export_from_r

Notes: Run this file in the R project, for the relative file paths to work. Uses R packages "foreign" and "isotone".

C2_2_estimation_param.do

Summary: Estimates the parametric selection function on various samples and returns residuals.

Inputs: Data/Temp/export_param

Outputs: Data/Temp/parametric_estimation

Notes: This script takes a minute or so to run.

C3_graph_basic.do

Summary: Plots the four baseline distributions of statistics and the corresponding subsamples.

Inputs: Data/Final/final_stars_supp

Outputs: Graphs stored in 'Figures' folder

Notes: Plots the panels of Figures 1 (sm_dist_cb; sm_dist_topJ), 5 (sm_dist_dataDriven_cb; sm_dist_explore_cb; sm_dist_noExpData_cb), 6 (sm_dist_control), 8 (sm_dist_minn; sm_dist_rba; sm_dist_rbnz), and 9 (sm_dist_pub_cb) in the accompanying paper. Charts formatted as per Brodeur *et al* (2016).

C3_graph_inputs.do

Summary: Plots the baseline and subsample distributions against inputs. The simu file inputs are outputs of Brodeur *et al* (2016).

Inputs: Data/Final/final_stars_supp

Data/Source/inputs/simu_wdi

Data/Source/inputs/simu_psid

Data/Source/inputs/simu_vhlss

Data/Source/inputs/simu_qog

Outputs: Graphs stored in 'Figures' folder

Notes: Graphs appear only in the online appendix. Charts formatted as per Brodeur *et al* (2016).

C3_graph_residuals.do

Summary: Plots selection functions and cumulated residuals.

Inputs: Data/Temp/export_from_r

Outputs: Graphs stored in 'Figures' folder

Notes: Plots the panels of Figure 7 (dissem_i_yf_cb_real; dissem_i_yf_cb_vhlss; dissem_i_yf_cb_qog) in the accompanying paper. Charts formatted as per Brodeur *et al* (2016).

C3_graph_residuals_param.do

Summary: Plots selection functions and cumulated residuals from parametric estimations.
Inputs: Data/Temp/parametric_estimation
Outputs: Graphs stored in 'Figures' folder
Notes: Graphs appear only in the online appendix. Charts formatted as per Brodeur *et al* (2016).

C3_summary_residuals.do

Summary: Retrieves the maximum of cumulated residuals and the corresponding t-value from parametric and non-parametric estimations.
Inputs: Data/Temp/parametric_estimation
Data/Temp/export_from_r
Outputs: Data/Final/summary_residuals

C3_summary_residuals_export.do

Summary: Creates the tables that give maximum cumulated residuals for various inputs and various subsamples
Inputs: Data/Final/summary_residuals
Outputs: Excel tables stored in 'Data/Final' folder
Notes: Must be run after C3_summary_residuals.do (see above). The output is the basis for Tables 2 (summary_residuals.xls) and 3 (summary_residuals_subsamples.xls, top and bottom sections) in the accompanying paper. summary_residuals_subsamples.xls is also used in the online appendix.

D0_master_and_create_outputs.R

Summary: Runs all of the p-curve analysis done in the other D scripts and creates the figures and table data that are shown in the paper and online appendix
Inputs: Data/Final/final_stars_supp.dta
Outputs: Graphs stored in 'Figures' folder
Notes: The script sources all of the p-curve analysis scripts and runs them in the correct order. Must be run after C3_summary_residuals.do (see above). The specific outputs produced are Figure 4, Figure A1 and Figure A2.

D1_construct_inputs.R

Summary: Prepares the sample and subsamples that the p-curve method are run on
Inputs: Data/Final/final_stars_supp.dta
Outputs: None
Notes: The main purpose of the script is to randomly select one test result from each paper. The random seed is pre-set to 1 to ensure that the same results are generated every time. If you do not run this line or choose a different random seed to may see slightly different results. The differences should be trivial.

D2_udf.R

Summary: Creates a set of bespoke (user-defined) functions that are used in the p-curve analysis.

Inputs: None

Outputs: None

Notes: Descriptions for what each function does and its purpose can be found in the script as comments.

D3_apply_method.R

Summary: Conducts the actual p-curve analysis on the sample and desired subsamples.

Inputs: Data/Final/final_stars_supp

Outputs: None

Notes: Must be run after D1_construct_inputs.R and D2_udf.R (see above) in the same R session. The results from this script are used to populate Table A1 in the online appendix and inform the text in the main paper. But no physical outputs are produced.

E1_create_z-curve_dissemination_bias_graphs.R

Summary: Creates Figure 7 in the format displayed in the main paper.

Inputs: Data/Temp/export_from_r

Outputs: Graph stored in 'Figures' folder

E1_create_z-curve_distribution_graphs.R

Summary: Creates Figures 1, 5, 6, 8 and 9 in the main paper.

Inputs: Data/Final/final_stars_supp.dta

Outputs: Graphs stored in 'Figures' folder

E1_hypothetical_p-curve_and_z-curve_distributions.R

Summary: Creates Figures 2 and 3 in the main paper.

Inputs: None

Outputs: Graphs stored in 'Figures' folder

starwars_graph_theme.R

Summary: Creates a bespoke ggplot theme for visually adjusting how the graphs in the main paper are formatted.

Inputs: None

Outputs: None

adjust_kernal_boundaries.R

Summary: Adjusts the boundary conditions used for the kernel estimates displayed in the z-curve distribution graphs.

Inputs: None

Outputs: None

Notes: The script was downloaded from github from the following address: https://github.com/echnasnovski/ggplot2/blob/geom_density-bounds/R/stat-density.r. It was not developed by us. It works by slightly adjusting a commonly used function in the 'stats' R package, which ggplot2 automatically calls on. The changes only affect graph aesthetics at the graph boundaries.

Reference list

Brodeur A, M Lé, M Sangnier and Y Zylberberg (2016), 'Star Wars: The Empirics Strike Back', *American Economic Journal: Applied Economics*, 8(1), pp 1–32.

Simonsohn U, LD Nelson and JP Simmons (2014), 'P-curve: A Key to the File-Drawer', *Journal of Experimental Psychology: General*, 143(2), pp 534–547.

11 February 2021