The Power of Tests for Detecting p-Hacking

Replication package

Graham Elliott
grelliott@ucsd.edu

Nikolay Kudrin
n.kudrin@queensu.ca

Kaspar Wüthrich
kwuthrich@ucsd.edu

This replication package contains the code for replicating all the results in the paper.

Software

- R version 4.3.2
- Matlab R2023b

Content

- The file 'BWfigures.m' generates figures for analytical examples (Figures 1-7 and 15-18) and saves them in the 'BW' folder.
- The file 'TwoSidedCurves.m' generates numerical analogs of figures 1-7 for 2-sided tests and saves them in the 'BW' folder.
- The file 'DataGeneration.m' generates initial Monte Carlo draws (it uses pveck2.m, pveckIV.m, BIC.m and NeweyWest.m functions; the descriptions of these functions are available in the corresponding m-files). The results are saved in the 'DGPs' folder.
- The file 'StataPart.do' generates signed t-statistics for calibrating Π to the RCT subsample of the data collected by (Brodeur, Cook, and Heyes, 2020)[1]. It creates 'Brodeur_data.csv'. Before running 'StataPart.do' download the folder with raw data from `https://doi.org/10.3886/E120246V1` (call this folder 'Data') and save it to the root folder. Then run 'StataPart.do' and save 'Brodeur_data.csv' to the root folder.
- The file 'dgps.m' generates the non-p-hacked and p-hacked distributions for various distributions of true effects (it uses NullAndAlt.m, NullAndAlt_var_bic.m, NullAndAlt_var_clust.m, NullAndAlt_iv_Fstat.m, PiGamma.m and EdgesN.m functions; the description of these functions is available in the corresponding m-files). The results are saved in the 'DGPs' folder.
- The file 'PowerFiguresSelected.m' generates the power curves reported in the paper. The results are saved in the 'PowerCurves' folder.

[1] Brodeur, Abel, Cook, Nikolai, and Heyes, Anthony. Data and Code for: Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics. Nashville, TN: American Economic Association [publisher], 2020. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2020-09-15. https://doi.org/10.3886/E120246V1

- The file 'MC_Tests.R' contains tests for p-hacking.
- The file 'MC_power.R' contains the function that calculates MC rejection rates for a given DGP.
- The file 'MC_power_main.R' replicates all Monte Carlo power curves reported in the paper.

To replicate all Monte Carlo results execute the files in the following order: 'DataGeneration.m' − >'dgps.m' (make sure to run 'StataPart.do' first)− >'MC_power_main.R' − >'PowerFiguresSelected.m'.

Disclaimer

This software is provided "as is" without warranty of any kind, expressed or implied. For questions and error reports, please contact the authors.