Information on the replication material that accompanies the paper ‘Bootstrap-inference for fixed-effect models’ by Ayden Higgins and Koen Jochmans

December 21, 2023

Overview

The code in this replication package allows a user to reproduce the empirical results and simulations in the paper. The folder ‘Replication’ contains the Matlab files `Table1.m` and `Table2.m`. These files reproduce the results in the respective tables. They both use data from the PSID, which are publicly available (details are given below). Table 1 can be reproduced in about 24 hours. Each design point in Table 2 equally takes about 24 hours to complete. Both tables contain an element of randomness induced through simulation. In `Table1.m` and `Table2.m` this randomness is controlled by working with fixed seeds when generating random numbers. This allows for exact replication. The folder ‘Test’ contains the same two Matlab files as above—only with the number of simulations and/or bootstrap replications set to a very small number—and the corresponding output files. These files can be useful to quickly verify that the code executes properly.

Data Availability and Provenance Statements

Statement about Rights

- ☒ I certify that the authors of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- ☒ I certify that the authors of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

License for Data

The data are licensed under a Creative Commons/CC-BY license.
See LICENSE.txt for details.

Summary of Availability

- ☒ All data **are** publicly available.
- ☐ Some data **cannot be made** publicly available.
- ☐ **No data can be made** publicly available.

Details on each Data Source

The PSID data used in this paper have been deposited by Ivan Fernández-Val in the ICPSR repository (https://doi.org/10.3886/E172521V1); see Fernández-Val (2022). The relevant file is `lfp_psid_fs.txt`.

Computational requirements

Software Requirements

Matlab with the optimization toolbox and parallel-computing toolbox.

The optimization toolbox is only used for the double bootstrap. The need for the parallel-computing toolbox can be removed by replacing the command `parfor` by `for` in the code (at the expense of a longer computing time).

The code was last run with version R2022b but older versions of Matlab should be fine.

Memory and Runtime Requirements

The code was run on a 6-core Intel-Xeon Mac Pro from 2016 with MacOS version 12.6.5.

For the replication files:

`Table1.m` reproduces the empirical results in Table 1 and takes about 24 hours to run.

`Table2.m` sequentially launches simulation exercises for various designs. This produces Table 2. Each design takes about 24 hours to complete, with some variation across the designs.

Run times can be lowered by adjusting the number of Monte Carlo replications (R), the number of (outer layer) bootstrap replications (B) or the number of inner bootstrap replications (C) for the double bootstrap. Table 1 in the paper was computed with $B = 9,999$ and $C = 999$ while Table 2 in the paper was generated with $R = 2,500$, $B = 999$, and $C = 0$ (Due to the computational burden, the iterated bootstrap was not computed in the Monte Carlo). These are the values to which these parameters are set in the files. The final output will, of course, not be fully invariant to changes in these parameters. We refer to Giacomini, Politis and White (2013) and Chang and Hall (2015) for the impact of this on the results of simulation experiments and on the performance of the double bootstrap, respectively.

For the test files: `Table1.m` has $B = 9$ and $C = 9$ while `Table1.m` runs with $B = 9$ and $R = 25$, $B = 9$, and $C = 0$. The test files run in a matter of minutes.

Description of programs/code

- `lfp_psid_fs.txt` contains the data and is called by both `Table1.m` and `Table2.m`.
- `Table1.m` reads the data and estimates the dynamic probit model. The results (point estimates, standard errors, and confidence intervals) for the autoregressive parameter as they appear in the paper are written to a text file called `Table1.txt`. The full set of estimation results (for all parameters; not reported in the paper) is written to the file `Table1_full.txt` for completeness.
- `Table2.m` performs a Monte Carlo experiment targeted to the empirical application for several values of the autoregressive parameter. It is essentially `Table1.m`, but has an extra layer built around it for the Monte Carlo. After estimating the model by maximum likelihood, it generates data from the fitted probit model (except that the autoregressive parameter takes the value given to it by the user, as dictated by the Monte Carlo design; here, 0, .5, 1, or 1.5) and then initiates a Monte Carlo simulation around that model. The results of the Monte Carlo are written in the same format as in the paper to `Table2.txt`.

Instructions to Replicators

1. Place the data set and files from the folder 'Replication' in your Matlab directory.

2. Run `Table1.m` to obtain the complete estimation results for the empirical application. These are printed to `Table1_full.txt`. Table 1 as reported in the paper is printed to `Table1.txt`.

3. Run `Table2.m` to replicate Table 2 in the paper.

The provided code reproduces:

- ☒ All numbers provided in text in the paper
- ☒ All tables and figures in the paper
- ☐ Selected tables and figures in the paper, as explained and justified below.

The bootstrap in Table 1 and the simulations in Table 2 have a component of Monte Carlo randomness to them. In the replication material this randomness is controlled by fixing the seed of the random number generator each time new data is generated (either for the bootstrap or for the Monte Carlo). This allows for exact reproducibility. If the aim is simply to test the code one can follow the steps above with the files `Table1.m` and `Table2.m` from the folder 'Test' rather than 'Replication'. The text files in the folder 'Test' contain the output that one should obtain on doing so.

References

Chang, J. and P. Hall (2015). Double-bootstrap methods that use a single double-bootstrap simulation. *Biometrika 102*, 203–214.

Fernández-Val, I. (2022). Datasets for Fernández-Val (2009), Journal of Econometrics 150(1), pp. 71-85. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2022-06-10. https://doi.org/10.3886/E172521V1.

Giacomini, R., D. N. Politis, and H. White (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory 29*, 567–589.

PSID. Panel Study of Income Dynamics, public use dataset. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI.