Description of the replication files of the empirical application in "Heterogeneous Autoregressions in Short T Panel Data", written by M. Hashem Pesaran and Liying Yang, 2024

June 22, 2024

The zipped file "PSID_application.py.jae.2024.zip" contains data files and associated R and MATLAB codes for the empirical application (earnings dynamics) results shown in Table 1 in Section 8 of the main paper and Tables S.19-S.23 in Section S.9 of the online supplement.

Estimators
==========

"FDAC" refers to the proposed estimator based on the autocorrelations of first differences. See Section 6 of the main paper.

"AAH", "AB", and "BB" refer to the estimators proposed by Chudik and Pesaran (2021), Arellano and Bond (1991), and Blundell and Bond (1998) for homogenous AR panels, respectively. "MSW" refers to a kernel-weighted likelihood estimator proposed by Mavroeidis et al. (2015) for heterogeneous AR panels. See Section 8.2 of the main paper for details.

Data
====

-- "PSID2.csv" contains a sample of households drawn from the Panel Study of Income Dynamics (PSID) data for the years 1976-1995, retrieved from the website: https://psidonline.isr.umich.edu. See "PSID2 1967-1996 data and notes.xlsx" for variable descriptions, sample selection criteria, and summary statistics.

-- "PCED2.csv" contains the GNP personal consumption expenditure deflator (using 1996 as the base year), retrieved from the Bureau of Economic Analysis of the US Department of Commerce in 2021.

-- The following data files contain the five- and ten-year sub-panels of the deflated PSID data for computing MSW estimates.
- "PSID2_educ_76_80.mat"
- "PSID2_educ_76_85.mat"
- "PSID2_educ_81_85.mat"
- "PSID2_educ_81_90.mat"
- "PSID2_educ_86_90.mat"
- "PSID2_educ_86_95.mat"
- "PSID2_educ_91_95.mat"

Programs
========

Compute FDAC and HomoGMM estimation results

-- "example_ar1.m" loads data from files "PSID2.csv" and "PCED2.csv", computes and saves the estimation results into "Hetro_AR_PSID_educ.mat". The sample is defined on lines 11-12. Codes are run separately for different sample choices.

Compute MSW estimation results

The MSW estimation can be executed on a local desktop, but it typically requires a considerable amount of time due to its lengthy duration — more than a day with no interruptions. Hence, we recommend executing the MC simulations on a cluster.

1. Generating results using a local desktop:

-- "Hetro_AR_PSID_msw_parallel.R" loads functions for computing the MSW estimator stored in "Hetro_AR_PSID_msw_functions.R" and loads data from "PSID2_educ_76_80.mat", "PSID2_educ_76_85.mat", "PSID2_educ_81_85.mat", "PSID2_educ_81_90.mat", "PSID2_educ_86_90.mat", "PSID2_educ_86_95.mat", and "PSID2_educ_91_95.mat". The estimation results are saved into "Hetro_AR_PSID_msw.RData".

2. Generating results using a cluster:

-- The script "Hetro_AR_PSID_msw.job" can be used to submit MC jobs to a cluster. Please adjust the parameters in the job file to match the cluster configuration, and install the packages used in the R script "Hetro_AR_PSID_msw_parallel.R" on the cluster before executing it. Both R scripts "Hetro_AR_PSID_msw_parallel.R" and "Hetro_AR_PSID_msw_functions.R" and the data files "PSID2_educ_76_80.mat", "PSID2_educ_76_85.mat", "PSID2_educ_81_85.mat", "PSID2_educ_81_90.mat", "PSID2_educ_86_90.mat", "PSID2_educ_86_95.mat", and "PSID2_educ_91_95.mat" should be uploaded to the path specified in the job script.

Tabulate the empirical estimation results

-- "Hetro_AR_PSID_tab.R" loads estimation results from "Hetro_AR_PSID_educ.mat" and "Hetro_AR_PSID_msw.RData" and generates Table 1 in Section 8 of the main paper and Tables S.19-S.23 in Section S.9 of the online supplement. The tables will be saved in "HetroAR_PSID_educ_results.xlsx".

Custom data sets
================

For estimation of panel AR(1) model using a custom data set please use "FDAC_custom_data_estimation.R". Please modify the code in line 12 of the R script to ensure

that the custom data set is correctly input into the estimation function. The estimation results will be saved in "FDAC_results.xlsx".

===
- The MATLAB script was written and executed using the MATLAB version R2022b.
- The R scripts were written and executed using the Rstudio version 2023.12.0+369.
- Before executing the MATLAB and R scripts, please ensure that the path is correctly set to the directory where data files and estimation results are stored, and the packages loaded in the R files are installed on either local desktops or clusters.
===

References

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. The Review of Economic Studies 58, 277-297.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115-143.

Chudik, A., and Pesaran, M. H. (2022). An augmented Anderson–Hsiao estimator for dynamic short-T panels. Econometric Reviews, 41, 416-447.

Mavroeidis, S., Sasaki, Y., and Welch, I. (2015). Estimation of heterogeneous autoregressive parameters with short panel data. Journal of Econometrics, 188, 219-235.