1 Replication materials for "Nonlinear household earnings dynamics, self-insurance, and welfare"

This readme file contains the information for the replication of the results provided in the paper. The code is structured in three folders:

1. **Earnings** folder. Contains all the information necessary to generate the nonlinear (NL) and canonical earnings processes from the PSID data.
2. **Processes** folder. This folder contains the Markov chains for the NL process obtained using the procedure described in the Earnings folder.
3. **Model** folder. This folder contains the code for our structural model, that takes the earlier earnings processes as inputs.

Please contact gonzalo.pardo.13@ucl.ac.uk if any difficulties arise using this material or if you have found a potential bug.

2 Earnings folder

2.1 PSID data

We use the publicly available PSID data, as described in Appendix A. Due to the PSID Conditions of Use, we are unable to directly provide the dataset, but interested users can download it free of charge from `https://psidonline.isr.umich.edu/` . Before loading the data to psid_transformation.R, we perform the following sample cleaning procedures and generate a panel.

1. Keep only households that are members of the SRC sample.
2. Link them over time, following the same household head.
3. Restrict the sample to 25-60 year-olds, and all waves before 1993.
4. Correct inconsistent ages. Whenever the age of a head of household does not evolve as expected in year t, verify if it can be reconstructed with information from $t-1$ and $t+1$. Otherwise, drop the observation.
5. Other standard variable cleaning (eliminate 9999 that represent don't knows, etc.).

2.2 Transforming PSID data

The code psid_transformation.R takes as input the cleaned PSID sample and shows:

1. The construction of asset income. All variables have been given intuitive names, and their precise relationship with PSID variables is available upon request.
2. The deflation of all monetary variables using the CPI-U to 2013 prices.
3. The construction of taxes, and disposable labor income.
4. The transformation of the sample to serve as an input to the Arellano, Blundell and Bonhomme (2017) procedure.
5. The computation of the variances of log earnings following Kaplan (2012) approach to remove time effects, as described in Appendix B, and also Figure B.1.
6. The generation of Figures 1 and C.1 (function incomedefs.R).

2.3 NL process

2.3.1 Applying the ABB method

The code required for the application of the Arellano et al. (2017) estimation is available on those authors' websites. At the time of submission it could be accessed here.

In ABB_coeffiCients.Xlsx ("Processes" folder) we report our estimated coefficients. They are estimated for 11 quantiles which are equally spaced (from 1/12=0.0833 to 11/12=0.9167).

The code provided by ABB generates Figure 2, 5b, 10, and D.4. The file abb_addition.m can be added to the ABB files to compute the statistics required for Figures 3 to 5.

The files eta_sample.txt and eps_sample.txt are the specific outputs from the ABB routine, which we use as an input to dfp_matrices.R

2.3.2 Discretizing the simulated sample

The file dfp_matrices.R allows to discretize any earnings sample, either after a transitory-permanent decomposition or directly from earnings data.

In the folder "Processes" we report the state spaces for the persistent component (eta_space), transitory component (eps_space), and the transition matrices for the persistent component (eta_tranmatrices), that represent the input to the structural model. The subfolder "Decompositions" includes the same state spaces and transition matrices for each of the decompositions discussed in Section 6.

2.4 Canonical process

The file sty_boot.m estimates the canonical earnings process by GMM and, via a bootstrap procedure, computes its standard errors. It takes as input the file psidsample_boot.csv, which is generated by psid_transformation.R and contains the relevant targeted variances (see point 5. in Section 2.2. above). This provides the information in Table 1.

2.5 Other elements and figures

The file varcons_variances_cex.R provides the code to compute the variances of log consumption following Kaplan (2012) approach to remove time effects, using the CEX data available in Heathcote, Perri and Violante (2010) (not provided).

The file more_graphs.R generates Figure 3(d), with all inputs provided, and Figures D.1. to D.3. and D.5.

3 Processes folder

This folder contains the crucial outputs from the "Earnings" section that serve as an input for the "Model" section. Users interested in the direct use of our provided state spaces and transition matrices can find related information inside the folder.

4 Model folder

The file main.m (and associated files in *mfiles*) contains the replication code for our structural model. It generates files with the key statistics for the graphs (in folder *Outputs*, provided) and simulations for the computation of the BPP coefficientes (also in folder Outputs, not provided to reduce size). All inputs for main.m are provided.

The file generate_graphs.m generates Figures 6, 7, 8, 9, D.6, D.7, and D.8 and Tables 3 and 4. All inputs are provided.

bpp_coefficients.R computes Blundell, Pistaferri and Preston (2008) insurance coefficients from a given earnings sample. It therefore generates Table 2 and Figure 11, and can be adjusted to generate Tables D.1. to D.3. It takes the simulations generated by main.m as input.

References

Arellano, Manuel, Blundell, Richard and Bonhomme, Stéphane (2017), 'Earnings and consumption dynamics: A non-linear panel data framework', *Econometrica* **85**(3), 693–734.

Blundell, Richard, Pistaferri, Luigi and Preston, Ian (2008), 'Consumption inequality and partial insurance', *The American Economic Review* pp. 1887–1921.

Heathcote, Jonathan, Perri, Fabrizio and Violante, Giovanni L (2010), 'Unequal we stand: An empirical analysis of economic inequality in the united states, 1967-2006', *Review of Economic dynamics* **13**(1), 15–51.

Kaplan, Greg (2012), 'Inequality and the life cycle', *Quantitative Economics* **3**(3), 471–525.