README

Replication package for "Optimal taxation and the Domar-Musgrave effect" by Brendan K. Beare and Alexis Akira Toda.

Overview

The code in this replication package reproduces the results in our published article using two data files. One data file contains historical U.S. capital income tax rates made publicly available by the Tax Foundation and the other contains historical U.S. household wealth shares made publicly available by Gabriel Zucman. Running the Matlab file `main.m` in the root directory reproduces Figures 1–12 in our published article, as well as the model-implied wealth shares reported in Table 2 in our published article. The replicator should expect the code to run for about 20 hours.

Data Availability and Provenance Statements

Data on historical U.S. capital income tax rates were downloaded from the Tax Foundation website. The maximum tax rate on capital gains may be downloaded from https://taxfoundation.org/data/all/federal/federal-capital-gains-tax-collections-historical-data (sixth column). The maximum tax rate on corporate income may be downloaded from taxfoundation.org/data/all/federal/historical-corporate-tax-rates-brackets (third column). The data are in the public domain.

Data on historical U.S. household wealth shares were downloaded from Gabriel Zucman's website: https://gabriel-zucman.eu/usdina. The data are contained in tabs `TE2b` and `TE2c` in the Excel file labeled "Tables II: distributional series". These data extend the series used in Saez and Zucman (2016). The data are in the public domain.

The empirical U.S. household wealth shares for 2001 reported in Table 2 in our published article are sourced from the working paper version of Davies et al. (2011) and from Saez and Zucman (2016), as discussed in Endnote 7 in our published article. The parameters used to calibrate our model are sourced from Cagetti and De Nardi (2006), Angeletos (2007), Gilchrist et al. (2009), Fagereng et al. (2020) and OECD (2023), as indicated in Table 1 in our published article. All of these data are in the public domain.

Statement of Rights

- ☑ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- ☐ I certify that the author(s) 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.

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 folder `data` in the root directory contains the data obtained from the Tax Foundation website and from Gabriel Zucman's website. The data from the Tax Foundation website are contained in the file `taxrates2023.xlsx`. The data from Gabriel Zucman's website are contained in the file `PSZ2020AppendixTablesII(Distrib).xlsx`.

The parameters used to calibrate our model, sourced from prior literature as indicated above and in Table 1 in our published article, are included in lines 4–26 of the Matlab file `calibration.m` in the root directory.

Dataset List

Data file	Source
`data/taxrates2023.xlsx`	https://taxfoundation.org/data
`data/PSZ2020AppendixTablesII(Distrib).xlsx`	https://gabriel-zucman.eu/usdina

Computational Requirements

Software Requirements

The code is run using Matlab, with the Optimization Toolbox and Parallel Computing Toolbox installed. The code was last run with Matlab Release 2024a.

Controlled Randomness

- ☐ Random seed is set at line _____ of program _____
- ☑ No Pseudo random generator is used in the analysis described here.

Memory, Runtime, Storage Requirements

Summary Approximate time needed to reproduce the analyses on a standard (2025) desktop machine:

- ☐ <10 minutes
- ☐ 10-60 minutes
- ☐ 1-2 hours
- ☐ 2-8 hours
- ☑ 8-24 hours
- ☐ 1-3 days

☐ 3-14 days
☐ > 14 days

Approximate storage space needed:

☐ < 25 MBytes

☐ 25 MB - 250 MB

☐ 250 MB - 2 GB

☐ 2 GB - 25 GB

☑ 25 GB - 250 GB

☐ > 250 GB

☐ Not feasible to run on a desktop machine, as described below.

Details The code was last run on an **8-core Intel-based desktop machine with Windows version 11 and 128GB of RAM**. Computation took **20 hours and 20 minutes**. Computations were run in parallel across 8 workers.

Description of programs/code

The Matlab file `main.m` located in the root directory generates all output. The root directory also contains 31 other Matlab files. These are all called from `main.m`, or from one another.

Further Matlab files are contained in the folders `CharFunTool-master`, `discretization-master` and `matlab-axis-label-alignment-master` located in the root directory. These do not need to be run manually, but are called when `main.m` is run.

- `CharFunTool-master` contains the Characteristic Functions Toolbox publicly available at https://github.com/witkovsky/CharFunTool. This toolbox provides algorithms for numerically inverting characteristic functions to obtain probability distributions, as described in Witkovský (2016). It is used to compute various probability distributions reported in our published article. The Characteristic Functions Toolbox is licensed to Viktor Witkovský under an MIT License. See `CharFunTool-master/LICENSE.txt` for details.
- `discretization-master` contains Matlab code for discretizing non-Gaussian Markov processes based on Farmer and Toda (2017) and publicly available at https://github.com/alexisakira/discretization. This toolbox is used to compute Markov state probabilities as described in Section 3.1 in our published article. It is licensed to Leland E. Farmer and Alexis Akira Toda under a GNU General Public License. See `discretization-master/LICENSE` for details.

- `matlab-axis-label-alignment-master` contains the Tools for Axis Label Alignment in MATLAB created by Ling Han and publicly available at https://github.com/phymhan/matlab-axis-label-alignment. This toolbox is used to improve the appearance of Figure 8 in our published article.

Instructions to Replicators

After saving the replication package to a local folder on your machine, right-click on this folder in the Matlab file navigation panel and select "Add to Path > Selected Folders and Subfolders". Then run the Matlab file `main.m`. This will generate all output files, saving them in the folder `results` located in the root directory. Note that the output files are included as part of our replication package. Running `main.m` will overwrite them.

Matlab may produce warning messages when `main.m` is run, indicating that figures are not generated correctly. Some "pop-up" images generated by the Matlab application may not display correctly. However, the images saved as PDFs in the folder `results` should display correctly.

It is not necessary to manually run any of the Matlab files contained in our replication package other than `main.m`. These files are all called from `main.m`, or from one another.

List of Generated Output

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.

All output files are saved in the folder `results` located in the root directory.

Figure/Table	Program	Line	Output file
Table 2	`main.m`	38	`table2.mat`
Figure 1	`main.m`	26	`fig_taxrates.pdf`
Figure 2(a)	`main.m`	38	`fig_tailProb.pdf`
Figure 2(b)	`main.m`	38	`fig_wealthShare.pdf`
Figure 3	`main.m`	42	`fig_alphaHat.pdf`
Figure 4(a)	`main.m`	46	`fig_tauK_tauL.pdf`
Figure 4(b)	`main.m`	46	`fig_tauK_W.pdf`
Figure 4(c)	`main.m`	46	`fig_tauK_AggK.pdf`
Figure 4(d)	`main.m`	46	`fig_tauK_AggC.pdf`
Figure 4(e)	`main.m`	46	`fig_tauK_R.pdf`
Figure 4(f)	`main.m`	46	`fig_tauK_omega.pdf`
Figure 5(a)	`main.m`	46	`fig_W_opt_noC_body.pdf`
Figure 5(b)	`main.m`	46	`fig_W_opt_noC_tail.pdf`

Figure/Table	Program	Line	Output file
Figure 6(a)	`main.m`	48	`fig_LK_gamma_tax.pdf`
Figure 6(b)	`main.m`	48	`fig_LK_gamma_prices.pdf`
Figure 7(a)	`main.m`	49	`fig_LK_sigma_tax.pdf`
Figure 7(b)	`main.m`	49	`fig_LK_sigma_prices.pdf`
Figure 8(a)	`main.m`	53	`fig_KLC.pdf`
Figure 8(b)	`main.m`	53	`fig_KLWnew.pdf`
Figure 8(c)	`main.m`	53	`fig_KLAggK.pdf`
Figure 8(d)	`main.m`	53	`fig_KLAggC.pdf`
Figure 8(e)	`main.m`	53	`fig_KLR.pdf`
Figure 8(f)	`main.m`	53	`fig_KLomega.pdf`
Figure 9(a)	`main.m`	53	`fig_W_opt_body.pdf`
Figure 9(b)	`main.m`	53	`fig_W_opt_tail.pdf`
Figure 10(a)	`main.m`	55	`fig_LKC_gamma_tax.pdf`
Figure 10(b)	`main.m`	55	`fig_LKC_gamma_prices.pdf`
Figure 11(a)	`main.m`	56	`fig_LKC_sigma_tax.pdf`
Figure 11(b)	`main.m`	56	`fig_LKC_sigma_prices.pdf`
Figure 12(a)	`main.m`	60	`fig_trans_R.pdf`
Figure 12(b)	`main.m`	60	`fig_trans_omega.pdf`
Figure 12(c)	`main.m`	60	`fig_trans_K.pdf`
Figure 12(d)	`main.m`	60	`fig_trans_C.pdf`
Figure 12(e)	`main.m`	60	`fig_trans_B.pdf`
Figure 12(f)	`main.m`	60	`fig_trans_revenue.pdf`

The output file `table2.mat` generated by line 38 in `main.m` produces the model-implied wealth shares reported in Table 2 in our published article. Some of the numbers produced must be converted to "bottom shares" from "top shares" by subtracting them from 100%.

Line 60 in `main.m` also generates the output file `votes.mat`. This file contains three numbers: the proportions of all agents, of all workers, and of all entrepreneurs who favor switching to the optimal tax regime, as discussed in the text of our published article.

References

Angeletos, G.-M. (2007) Uninsured idiosyncratic investment risk and aggregate saving. *Review of Economic Dynamics*, 10(1), 1–30. Available from: https://doi.org/10.1016/j.red.2006.11.001

Cagetti, M. & De Nardi, M. (2006) Entrepreneurship, frictions and wealth. *Journal of Political Economy*, 114(5), 835–870. Available from: https://doi.org/10.1086/508032

Fagereng, A., Guiso, L., Malacrino, D. & Pistaferri, L. (2020) Heterogeneity and persistence in returns to wealth. *Econometrica*, 88(1), 115–170. Available from:

https://doi.org/10.3982/ECTA14835

Farmer, L.E. & Toda, A.A. (2017). Discretizing nonlinear, non-Gaussian Markov processes with exact conditional moments. *Quantitative Economics*, 8(2), 651–683. Available from: https://doi.org/10.3982/QE737.

Gilchrist, S., Yankov, V. & Zakrajšek, E. (2009) Credit market shocks and economic fluctuations: Evidence from corporate bond and stock market. *Journal of Monetary Economics*, 56(4), 471–493. Available from: https://doi.org/10.1016/j.jmoneco.2009.03.017

OECD (2023). *Taxing Wages 2023: Indexation of Labour Taxation and Benefits in OECD Countries*. OECD Publishing. Available from: https://doi.org/10.1787/8c99fa4d-en

Saez, E. & Zucman, G. (2016) Wealth inequality in the United States since 1913: Evidence from capitalized income tax data. *Quarterly Journal of Economics*, 131(2), 519–572. Available from: https://doi.org/10.1093/qje/qjw004

Toda, A.A. & Wang, Y. (2021) Efficient minimum distance estimation of Pareto exponent from top income shares. *Journal of Applied Econometrics*, 36(2), 228–243. Available from: https://doi.org/10.1002/jae.2788

Witkovský, V. (2016) Numerical inversion of a characteristic function: An alternative tool to form the probability distribution of output quantity in linear measurement models. *Acta IMEKO*, 5(3), 32–44. Available from: https://doi.org/10.21014/acta_imeko.v5i3.382