

Health Sector Structural Change

Replication Package “README” File

Nick Pretnar and Maria Feldman

October 29, 2025

Overview

The code and data in this replication package construct all analyses files, tables, and figures in the main draft and supplemental appendix of the paper using Bash shell scripts, R, and Fortran. Four files in `main_code`, run sequentially, can alone compute all of the figures and tables in the main draft and supplemental appendix: 1) `main_code/0_setup.R`, 2) `main_code/1_data_work.R`, 3) `main_code/2_simulate_model.sh`; 4) `main_code/4_parameter_plots.R`. The file `main_code/2_estimate_model.sh` runs the full Monte Carlo simulation estimator from scratch, but this is not needed in order to produce the figures and tables in the text, since the `data/analysis` sub-directories contain the parameter estimates from this file. NOTE: the full estimator will require 10 threads in parallel (cores) and take over 3 days to run all of the different calibrations. The file in the main directory `main.sh` is a shell script that will run all of the dependent scripts in order of required execution, prompting the user about half way through to press “1” if the full estimator is desired or “0” if only the simulated estimator that spits out the figures and tables should be run. This file needs to be made executable by executing the following command in a bash shell pointed to the directory wherein resides the file `main.sh` prior to running it: `chmod u+x main.sh`. Then run, `bash main.sh`, enter “1” or “0” according to the prompt, and the results shall be had.

Finally, all data used to produce descriptive statistics and model-dependent results are publicly available with sources described in detail below.

Data Availability and Provenance Statements

This paper uses several different datasets. For aggregate personal consumption expenditure, gross domestic product, and labor inputs by sector, we turn to the National Income and Product Accounts from the U.S. Bureau of Economic Analysis (BEA). For wages by sector we turn to the U.S. Bureau of Labor Statistics (BLS). For health-sector markups we at times use estimates from Figures 6 and 7 of Horenstein and Santos (2019), though we also estimate

relative markups directly from input/output data, independent of their estimates. For non-health-sector total factor productivity (TFP), we use TFP as presented in the Penn World Tables 10.0.1; we also use Penn World Tables 10.0.1 data for the time-varying non-health-sector labor share of income (Feenstra, Inklaar, and Timmer 2015). For capital input data by sector, we use the BEA’s Fixed Asset Tables. Population by year and age come from the United Nation’s Department of Economic and Social Affairs, Population Division. Accidental mortality data come from Health, United States 2017 from the National Center for Health Statistics.

All raw data used, prior to pre-processing by the script `main_code/1_data_work.R`, are contained in `data/raw`. Data descriptions and sources, including formal references are presented below.

Aggregate Data from the National Income and Product Accounts (NIPA)

All NIPA data we use are included in the replication package download. Tables 2.5.3, 2.5.4, and 2.5.5 were originally accessed on November 11, 2023; these tables are saved as `data/raw/bea_nipa_2_5_3.csv`, `data/raw/bea_nipa_2_5_4.csv`, and `data/raw/bea_nipa_2_5_5.csv`. To access these datasets independent of those we have provided, navigate to <https://www.bea.gov/itable/national-gdp-and-personal-income> and use the “Interactive Data Tables” feature, “Section 2 – Personal Income and Outlays” (BEA 2023a; BEA 2023b; BEA 2023c). Tables 2.5.3, 2.5.4, and 2.5.5 are available annually and should be modified to include years 1948-2022. Table 1.5.5 was originally accessed on July 24, 2024, and is saved as `data/raw/bea_nipa_1_5_5.csv`. To access this table independent of that which we have provided, navigate to <https://www.bea.gov/itable/national-gdp-and-personal-income> and use the “Interactive Data Tables” feature, “Section 1 – Domestic Product and Income” (BEA 2024b). Table 1.5.5 is available both quarterly and annually: the annual option should be selected along with years 1948-2022. Table 3.15.5 was originally accessed on July 24, 2024, and is saved as `data/raw/bea_nipa_3_15_5.csv` (BEA 2023d). To access this table independent of that which we have provided, navigate to <https://www.bea.gov/itable/national-gdp-and-personal-income> and use the “Interactive Data Tables” feature, “Section 3 – Government Current Receipts and Expenditures.” Table 3.15.5 is available annually: years 1959-2022 should be selected.

NIPA Tables 6.5B, 6.5C, and 6.5D measure full-time equivalent employees by industry (BEA 2020b; BEA 2020c; BEA 2020d). Table 6.5B ranges from 1948-1987, Table 6.5C from 1987-2000, and Table 6.5D from 1998 to present, with industry re-definitions occurring throughout. These datasets were accessed on September 7, 2020, and are saved as `data/raw/nipa_table_6_5B.csv`, `data/raw/nipa_table_6_5C.csv`, and `data/raw/nipa_table_6_5D.csv`.

Finally, we use annualized real GDP data for model validation purposes only.

This is “GDPC1” from FRED (BEA 2024a). It is saved as `data/raw/gdp.Rdata`.

Wages by Sector from the BLS

This data is only used for validation in the supplemental appendix. We select education and health sector wages from the BLS via FRED at the St. Louis Fed, series ID “CES6500000003” (BLS 2024a). This data is contained in `data/raw/educ_health_wages.csv`. We also select economy-wide wages from series ID “CES0500000003” (BLS 2024b). This data is contained in `data/raw/total_priv_wages.csv`.

Penn World Tables, 10.0.1 (Feenstra, Inklaar, and Timmer 2015)

Two data series from the Penn World Tables are used — total factor productivity estimates, accessed from FRED with series code “RTFPNAUSA632NRUG,” and time-varying aggregate labor share estimates, accessed from FRED with series code “LABSHPUSA156NRUG” (University of Groningen and University of California, Davis 2019b; University of Groningen and University of California, Davis 2019a). These are saved as `data/raw/tfp.csv` and `data/raw/labor_share.csv`, respectively.

Capital Inputs by Sector from the BEA’s Fixed Asset Tables

For capital by sector we use BEA Fixed Asset Tables 3.1ESI and 3.2ESI. Table 3.1ESI measures the “Current-Cost Net Stock of Private Fixed Assets by Industry,” and Table 3.2ESI measures the “Chain-Type Quantity Indexes for Net Stock of Private Fixed Assets by Industry” which we use to construct real quantities of capital by industry (BEA 2020a; BEA 2022). Fixed Asset tables can be accessed by navigating to <https://www.bea.gov/itable/fixed-assets> and clicking “Interactive Data Tables” then downloading tables in “Section 3 – Private Fixed Assets by Industry.” These tables are saved as `data/raw/bea_fixed_assets_3_1.csv` and `data/raw/bea_fixed_assets_3_2_ESI.csv`.

Population Data by Age from the United Nations (UN)

Population data by age and U.S. life tables come from the UN’s World Population Prospects from the Department of Economic and Social Affairs, Population Division. To access such data, navigate here <https://population.un.org/wpp/downloads?folder=Standard%20Projections&group=Population>, selecting the “Standard Projections” and “Population” buttons, then download the “Population by Single Age – Both Sexes (XLSX)” file for single age population and “Abridged Life Table – Both Sexes (XLSX)” for the life tables. For the median age data, navigate to “Special Aggregates” and download “Most used data (XLSX)” which contains the median age statistic. We include population by age, life tables, and median age statistics in three files: `data/raw/us_pop_data.csv`, `data/raw/us_life_tables.csv`, and `data/raw/us_median_age.csv` (UN 2021c; UN 2021a; UN 2021b).

Accidental Mortality Rates by Age from the U.S. Centers for Disease Control and Prevention (CDC)

Accidental mortality rates come from the “Health, United States, 2017” technical report from the CDC, which can be accessed as an archival version here: <https://www.cdc.gov/nchs/hus/contents2017.htm> (National Center for Health Statistics 2017). Accidental mortality data by age can be computed conditional upon accident type. These data are exogenous to our model and taken as given and fixed in our estimation routines. You will find all the necessary datasets and a PDF file with their descriptions in `data/raw/HEALTH_2017_accidental_deaths`. Table 021 and 027 through 032 are used.

Statement about Rights

- ☑ 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.

License for Data

All data are publicly available and no license is required.

Details on each Data Source

All data listed below are provided and contained in `data/raw`:

Computational requirements

The analyses use a combination of R and Fortran with commands wrapped in several, executable bash shell scripts. `main.sh` is the main file which should be made executable and run from its directory. With this file, you will be prompted to enter “1” or “0” after some preliminary calculations. Entering “1” will force you to re-run the full estimator, `main_code/2_estimate_model.sh`, while entering “0” will only run the scripts that generate all tables and figures from the provided parameters which have already been estimated — `main_code/2_simulate_model.sh`. Note that 10 cores (or threads) which can run independently in parallel are required for the full estimator, `main_code/2_estimate_model.sh`, while only 1 core is required to run `main_code/2_simulate_model.sh` and all of the R scripts.

NOTE: parallelization is only required if fully re-estimating the model. If choosing “1” to fully re-estimate the model, then if less cores are required for parallelization due to system constraints, the replicator should do the following:

1. Open `main_code/2_estimate_model.sh`.
2. Comment out or delete enough “seeds” entries, such that the length of the new “seeds” vector corresponds to the number of parallel cores you are choosing to operate.

3. *This MAY change the estimation results and make them less accurate. To replicate the full estimator, all parallel chains with all corresponding starting seeds should be run prior to executing `analysis/$full_path/2_monte_carlo_analysis.R`.*

Software Requirements

The following software/programs should be installed BEFORE executing the code:

- R (version 4.4.2 “Pile of Leaves” should be used). Packages required are as follows (downloaded as of 2025-07-30; note that the file `main_code/0_setup.R` automatically ensures the following packages and their correct versions are installed):
 - `readr`; `reshape2`; `tidyr`; `scales`; `Hmisc`; `readxl`; `EnvStats`; `ivreg`; `xtable`; `stringr`; `dplyr`
- `gfortran` of the GNU Fortran/GCC compiler set (version Homebrew GCC 14.2.0 used).

The following computational libraries are also required, but should be already tightly integrated into the MacOS eco-system. It is decidedly *not* recommended to separately install these programs using Homebrew (for example), as the binaries included in these programs are integral to the entire MacOS operating system. If, though, manual installation is for some reason necessary, you can install BLAS with Homebrew: `brew install openblas`. You can also install lapack with Homebrew: `brew install lapack`. Again, this is *not* recommended.

- BLAS (Basic Linear Algebra Subprograms)
- LAPACK (Linear Algebra PACKage)

Controlled Randomness

- ☒ Random seed is passed to underlying Fortran programs at line 60 of program `main_code/2_estimate_model.sh` (NOTE: this is only relevant if running the full estimator, selecting “1” when prompted: `main_code/2_estimate_model.sh`).
- ☐ No Pseudo random generator is used in the analysis described here.

Memory, Runtime, Storage Requirements

Summary Computational requirements depend on whether or not you are re-running the full estimators/calibrations from scratch (i.e., if you select “1” when prompted by the `main.sh` shell script).

Approximate time needed to reproduce the analyses given estimated parameters (i.e., selecting “0” when prompted, only to generate tables and figures) on a standard (CURRENT YEAR) 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 time needed to estimate the full model and generate all counterfactual tables and figures (i.e., selecting “1” when prompted) on a standard (CURRENT YEAR) 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 a **10-core MacBook Pro model 18,1 with Apple M1 Pro chip running MacOS Sequoia 15.5**. Total computation time when running the full estimator (i.e., selecting “1” when prompted) was 74 hours. Total computation time when just reproducing tables and figures from given parameters (i.e., selecting “0” when prompted) was < 2 minutes.

Description of programs and code

The program `main.sh` is a shell script that envelops all other programs and should be made executable, `chmod u+x main.sh`, then executed with `bash main.sh` or `./main.sh`. In what follows below, we describe the main container scripts that execute the model code. Note that the code to solve, estimate, and simulate the full, general equilibrium models under different parametric calibrations is

contained in the sub-directories of `data/analysis`, which we discuss in further detail below.

The program `main.sh` executes the following container scripts, sequentially, in order:

- `main_code/0_setup.R` sets up the R environment, ensuring that all packages are the correct version (NOTE: R's binaries should be installed prior to running this code).
- `main_code/1_data_work.R` takes in the raw data in `data/raw` and processes the data to compute data moments targeted by the model and several summary statistic tables and figures in both the main text and the supplemental appendix. Specifically, `main_code/1_data_work.R` computes the following figures in the main text: `figure1.png`, `figure2.png`, `figure3.png`, and `figure4.png`. It also writes model moments to be targeted in the structural estimation to the various sub-directories in the `data/analysis` directory where calibrations take place. We discuss those code files below.
- `main_code/2_estimate_model.sh` runs the entire, structural estimation routine, compiling the Fortran programs in the sub-directories of `data/analysis/` and then executing the estimator, which writes out the estimation results into the corresponding sub-directories. This file is executed if the user inputs "1" when prompted by `./main.sh`.
- `main_code/2_simulate_model.sh` only runs the programs necessary to subsequently generate all of the tables and figures post-estimation.
- `main_code/3_parameter_plots.R` generates all tables and figures resulting from the full, general equilibrium model estimation. Specifically, in the main text the code generates `figure5.png`, `figure6.png`, `figure7.png`, and `figure8.png`, as well as `table2.tex` and `table3.tex`. Note that Table 1 is hard-coded directly into the `.tex` file.

The full, general equilibrium model solution, estimation, and simulation scripts are included in each of the following sub-directories of `data/analysis`, with each directory corresponding to a different calibration strategy (involving different assumptions regarding markups and α_c parameters, as discussed in Section III of the main text (numbering corresponds to the calibration "numbers" used in the first paragraph of Section III): 1) `GE_Calibration_Mu_Growth_Past_2015`, 2) `GE_Calibration_Constant_Alpha_c`, 3) `GE_Calibration_Variable_Alpha_c`, 4) `GE_Calibration_HS_Markups_Full_Calibration`, and 5) `GE_Calibration_No_Mu_Growth`. Note that the sub-directory `GE_Calibration_HS_Markups_Sim_Only` merely simulates a model using parameters calibrated in `GE_Calibration_Constant_Alpha_c` but with the Horenstein and Santos (2019) relative markups. Results associated with simulations in `GE_Calibration_HS_Markups_Sim_Only` are entirely restricted to the Supplemental Appendix.

Each sub-directory, 1) `GE_Calibration_Mu_Growth_Past_2015`, 2) `GE_Calibration_Constant_Alpha_c`, 3) `GE_Calibration_Variable_Alpha_c`, 4) `GE_Calibration_HS_Markups_Full_Calibration`, and 5) `GE_Calibration_No_Mu_Growth`, contains the following set of Fortran

.f90 scripts, the purpose of each we now describe:

- `MAIN.f90` is the primary wrapper program that uses and encapsulates all of the other programs in the Fortran set. The program (after compilation) takes in command-line arguments that are passed to it from either of the two wrappers, `main_code/2_estimate_model.sh` or `main_code/2_simulate_model.sh`. The parameters passed via the command-line are logicals which tell the compiled program to either estimate the full model via the Monte Carlo Simulated Method of Moments (SMM) or to merely simulate the equilibrium path given the estimated parameters contained in the file `parms_min.txt`, of which there exists exactly one per sub-directory.
- `kind_module.f90` declares the double-precision data type.
- `Global_Values.f90` sets Fortran parameters and storage objects as global.
- `toolbox_standalone.f90` and `toolbox_outer.f90` are variations of toolboxes from Fehr and Kindermann (2018) that are copied in different files with different function names so they can be called within different sub-routines.
- `prob.f90` contains code to generate random variates used in the SMM.
- `solver.f90` contains the code to solve the model, including solving the household's problem using endogenous grid methods and solving for the general equilibrium objects (prices) following the procedure in Krueger and Ludwig (2007).
- `SMM.f90` contains the code to read in data moments and compute model moments under different parameters drawn using a Monte Carlo routine. It wraps the execution of `solver.f90`, which solve the model, including the transition paths, conditional upon the parameters fed to it by `SMM.f90`.

List of Tables and Figures

The following tells where tables and figures are produced within the replication package. (NOTE: one cannot necessarily expect to execute individual scripts/programs in isolation, outside of the chain of execution summarized by the shell script, “main.sh” without compromising the reproducible integrity of the tables and figures.) We list these figure by figure (table by table) as they appear in the text:

- Figure 1, `main_code/1_data_work.R`
- Figure 2, `main_code/1_data_work.R`
- Figure 3, `main_code/1_data_work.R`
- Figure 4, `main_code/1_data_work.R`
- Table 1 is hard coded into the main manuscript tex file

- Table 2, `main_code/3_parameter_plots.R`
- Figure 5, `data/analysis/GE_Calibration_Mu_Growth_Past_2015/2_counterfactual_analyses.R`
- Figure 6, `main_code/3_parameter_plots.R`
- Figure 7, `data/analysis/GE_Calibration_Mu_Growth_Past_2015/2_counterfactual_analyses.R`
- Figure 8, `data/analysis/GE_Calibration_Mu_Growth_Past_2015/2_counterfactual_analyses.R`
- Table 3, `data/analysis/GE_Calibration_Mu_Growth_Past_2015/2_counterfactual_analyses.R`

Instructions to Replicators

All raw data is included in the replication package and need not be preliminarily downloaded prior to execution.

- Download the replication package called `replication_package_pretnar_feldman_2025` and, in a Terminal window or similar bash emulator, navigate to the replication package’s directory using `cd $PATH` where `$PATH` should be replaced by the actual path of the directory.
- Make the main script executable by running `chmod u+x main.sh`.
- Execute the main script using `bash main.sh` or `./main.sh` and follow the prompts accordingly, when requested to press “1” to run full estimator or “0” for mere simulation.

NOTE: replicators may encounter reasonably very small deviations from the computational outcomes in the paper, due to variations in machine architecture, compiler’s random number generators, different compiler installation methods, Xcode versions (if using a Mac), etc.

References

- BEA (2020a). *Fixed Assets Accounts, Table 3.1ESI (version September 2, 2020)*. Accessed: 2022-05-03. URL: <https://www.bea.gov/itable/fixed-assets> (cit. on p. 3).
- (2020b). *Table 6.5B. Full-time Equivalent Employees by Industry (version July 31, 2020)*. Accessed: 2020-09-07. URL: <https://www.bea.gov/itable/national-gdp-and-personal-income> (cit. on p. 2).
- (2020c). *Table 6.5C. Full-time Equivalent Employees by Industry (version July 31, 2020)*. Accessed: 2020-09-07. URL: <https://www.bea.gov/itable/national-gdp-and-personal-income> (cit. on p. 2).
- (2020d). *Table 6.5D. Full-time Equivalent Employees by Industry (version July 31, 2020)*. Accessed: 2020-09-07. URL: <https://www.bea.gov/itable/national-gdp-and-personal-income> (cit. on p. 2).
- (2022). *Fixed Assets Accounts, Table 3.2ESI (version September 30, 2022)*. Accessed: 2023-05-15. URL: <https://www.bea.gov/itable/fixed-assets> (cit. on p. 3).

- BEA (2023a). *Table 2.5.3. Real Personal Consumption Expenditures by Function, Quantity Indexes (version September 29, 2023)*. Accessed: 2023-11-11. URL: <https://www.bea.gov/itable/national-gdp-and-personal-income> (cit. on p. 2).
- (2023b). *Table 2.5.4. Price Indexes for Personal Consumption Expenditures by Function (version September 29, 2023)*. Accessed: 2023-11-11. URL: <https://www.bea.gov/itable/national-gdp-and-personal-income> (cit. on p. 2).
- (2023c). *Table 2.5.5. Personal Consumption Expenditures by Function (version September 29, 2023)*. Accessed: 2023-11-11. URL: <https://www.bea.gov/itable/national-gdp-and-personal-income> (cit. on p. 2).
- (2023d). *Table 3.15.5. Government Consumption Expenditures and Gross Investment (version November 17, 2023)*. Annual Version, Accessed: 2024-07-24. URL: <https://www.bea.gov/itable/national-gdp-and-personal-income> (cit. on p. 2).
- (2024a). *Real Gross Domestic Produce [GDPC1]*. Retrieved from FRED, Federal Reserve Bank of St. Louis. Accessed December 2, 2024. URL: <https://fred.stlouisfed.org/series/GDPC1> (cit. on p. 3).
- (2024b). *Table 1.5.5. Gross Domestic Product, Expanded Detail (version June 27, 2024)*. Annual Version, Accessed: 2024-07-24. URL: <https://www.bea.gov/itable/national-gdp-and-personal-income> (cit. on p. 2).
- BLS (2024a). *Average Hourly Earnings of All Employees, Private Education and Health Services [CES6500000003]*. Retrieved from FRED, Federal Reserve Bank of St. Louis. Accessed December 13, 2024. URL: <https://fred.stlouisfed.org/series/CES6500000003> (cit. on p. 3).
- (2024b). *Average Hourly Earnings of All Employees, Total Private [CES0500000003]*. Retrieved from FRED, Federal Reserve Bank of St. Louis. Accessed December 13, 2024. URL: <https://fred.stlouisfed.org/series/CES0500000003> (cit. on p. 3).
- Feenstra, Robert, Robert Inklaar, and Marcel Timmer (2015). “The Next Generation of the Penn World Table”. *American Economic Review* 105.10. Available for download at www.ggd.net/pwt, pp. 3150–3182 (cit. on pp. 2, 3).
- Fehr, Hans and Fabian Kindermann (2018). *Introduction to Computational Economics using Fortran*. Oxford: Oxford University Press (cit. on p. 8).
- Horenstein, Alex R. and Manuel S. Santos (2019). “Understanding Growth Patterns in US Health Care Expenditures”. *Journal of the European Economic Association* 17.1, pp. 284–326 (cit. on pp. 1, 7).
- Krueger, Dirk and Alexander Ludwig (2007). “On the consequences of demographic change for rates of returns to capital, and the distribution of wealth and welfare”. *Journal of Monetary Economics* 54, pp. 49–87 (cit. on p. 8).
- National Center for Health Statistics (2017). *Health, United States — Data Finder: 2017 edition*. Hyattsville, MD. Available from: <https://www.cdc.gov/nchs/health/data-finder.htm?year=2017>. Accessed June 16, 2021 (cit. on p. 4).
- UN (2021a). *World Population Prospects, Abridged Life Table – Both Sexes (XLSX) (version June 3, 2021)*. Accessed: 2021-06-03. URL: <https://populati>

- on.un.org/wpp/downloads?folder=Standard\%20Projections\&group=Population (cit. on p. 3).
- UN (2021b). *World Population Prospects, Median by Age (XLSX) (version June 3, 2021)*. Accessed: 2021-06-03. URL: <https://population.un.org/wpp/downloads?folder=Standard\%20Projections\&group=Population> (cit. on p. 3).
- (2021c). *World Population Prospects, Population by Single Age – Both Sexes (XLSX) (version June 3, 2021)*. Accessed: 2021-06-03. URL: <https://population.un.org/wpp/downloads?folder=Standard\%20Projections\&group=Population> (cit. on p. 3).
- University of Groningen and University of California, Davis (2019a). *Share of Labour Compensation in GDP at Current National Prices for United States [LABSHPUSA156NRUG]*. Retrieved from FRED, Federal Reserve Bank of St. Louis. Accessed December 13, 2024. URL: <https://fred.stlouisfed.org/series/LABSHPUSA156NRUG> (cit. on p. 3).
- (2019b). *Total Factor Productivity at Constant National Prices for United States [RTFPNAUSA632NRUG]*. Retrieved from FRED, Federal Reserve Bank of St. Louis. Accessed December 19, 2022. URL: <https://fred.stlouisfed.org/series/RTFPNAUSA632NRUG> (cit. on p. 3).