README File for "Insurance, Redistribution and the Inequality of Lifetime Income"

November 15, 2024

1 Overview

This README file accompanies the replication package for the paper titled "Insurance, Redistribution and the Inequality of Lifetime Income." The code in this replication package reproduces the results in the paper and online appendix using a mixture of Stata and Matlab. All cross-references to figures and tables numbers refer to the manuscript and web appendix dated November 15, 2024. Transfer this directory, along with all its subdirectories, onto your machine. Please be aware that you might come across some directories that are empty. Even so, go ahead and transfer them, as these will be filled with necessary data as you execute the programs. Replication is likely to take several days, depending on computational resources.

2 Data Availability and Provenance Statements

This study uses survey data from the German Socio-Economic Panel (SOEP) that we cannot share.[1] To reproduce the results in the paper three SOEP versions are required:

1. SOEP – Version 27:
   SOEP (2011). Socio-Economic Panel (SOEP), data for years 1984-2010, version 27. doi:10.5684/soep.v27
2. SOEP – Version 33:
   SOEP (2017). Socio-Economic Panel (SOEP), data for years 1984-2016, version 33. doi:10.5684/soep.v33
3. SOEP – Version 35:
   SOEP (2019). Socio-Economic Panel (SOEP), data for years 1984-2018, version 35. doi:10.5684/soep.v35

- Access requires submitting an application to the German Institute for Economic Research (DIW Berlin). Please follow this link for instructions on accessing the SOEP data: `https://www.diw.de/en/diw_01.c.678568.en/research_data_center_soep.html`
- Please use the SOEP versions referenced above.
- Save the original SOEP data versions in the respective folders in `Data/SOEP_confid/SOEPvXX`.

Moreover, this study uses the following publicly accessible data:

1. Lifetables from Human Mortality Database
   - **HMD** (2024). Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at www.mortality.org (data downloaded on April 15, 2024)
   - We provide this data in the file `Data/HMD/mltper_1x1.csv`.

[1] Wagner et al. (2007) and Goebel et al. (2019) describe the SOEP.

Statement about Rights

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

3 Computational Requirements

Most parts of this replication package can be run on a desktop machine (including data cleaning, estimation and post-estimation analyses), although some of the procedures may take several days to run. Meanwhile, it is more practical to run the estimation processes itself on a computing cluster. Estimation and simulation procedures in Matlab allow for parallel processing. Details on how to specify the parallel computing environment, the time required to run each part of the replication and further particulars about the computing resources are provided below in Section 5.

4 Overview of the Repository

Structure of repository:

- **Data:**
  This directory contains all data sources as well as all datasets that are generated during the data processing, simulation and analysis procedures. Please note that all files in directory `Data/SOEP_confid/` cannot be redistributed.
- **Stata code:**
  This directory contains all Stata do-files that are used to prepare the data, estimate auxiliary models and conduct the post-estimation analysis. All required Stata packages are installed to this directory. All figures of the manuscript and web appendix are saved to folders `StataCode/02_figures/`, and all numerical values and tables are written to appropriately labeled Sheets of Excel book `StataCode/03_tables/CollectedResults.xlsx`.
- **Matlab code:**
  This directory contains all Matlab functions that are used to estimate the life-cycle model and to simulate the scenarios. Separate folders for all generated input and output files are included.

Execution order:
Due to the use of Stata and Matlab, and given the varying computational requirements, the code execution is separated into four blocks. Associated Master scripts are saved in the root directory of the repository. Beyond the preservation of the execution order specified below, blocks can be run independently of each other.

1. **Data preparation and auxiliary estimations (Stata)**
   Executable using script `Master_Part1_StataDataPrep.do`.
   Extracts and prepares all required information from data sources. Generates the main estimation sample and all additional datasets required in auxiliary regressions. Generates all required inputs for the estimation of the life-cycle model and executes auxiliary regressions (heterogeneous health shock profiles, mortality risk profiles and involuntary separation risk profiles).
2. **Model estimation (Matlab)**
   Executable using script `Master_Part2_MatlabEstim.m`.
   Runs all estimation procedures of the main life-cycle model for the main calibration and robustness checks using alternative specifications of the calibrated preference parameters.
3. **Simulation exercises (Matlab)**
   Executable using script `Master_Part3_MatlabSim.m`.
   Employs the estimated life-cycle model to construct simulated samples for the baseline and counterfactual environments. Computes welfare effects for the policy reform scenario. Runs the simulations for robustness checks to alternative calibrations of preference parameters.
4. **Post-estimation analysis (Stata)**
   Executable using script `Master_Part4_StataAnalysis.do`.
   Runs the analysis on all simulated scenarios and generates results presented in the manuscript.

Generated figures and tables:
A list of all figures and tables presented in the manuscript and web appendix, with reference to the specific script that generates each, is provided in Table 1 and Table 2 at the end of this document. Labelling of figures and tables is in reference to the manuscript and web appendix dated November 15, 2024.

- All generated figures are saved to `StataCode/02_figures`.
- All numerical values and tables are saved to the Excel book `StataCode/03_tables/CollectedResults.xlsx`.

Controlled Randomness

The simulations use pseudorandom number generators. For reproducibility purposes, deterministic seeds are used. A random seed is set at:

- line 88 of `calibrate.m`; used in line 30 of `drawgen.m`.
- line 314 of `analysis_dataprep.do`
- line 46 and 182 of `rob_partime.do` (must be identical)
- line 134 of `proginc.do`

5 Instructions to Replicators

The steps in this section must be executed in the order listed.

All required Stata packages are installed to the repository on execution of the Master script of Part 1: `estout`, `ineqdeco` and `ineqdec0`.

At least 35 GB of hard disk space must be available to save the SOEP datasets and simulated data samples.

5.1 Part 1: Primary data work and auxiliary estimations

Notes:

- Computations described in this section were implemented using Stata version 17 (SE-Standard Edition) on a `MacBook Pro M1 2021 16GB` machine.

Steps:

1. Follow the instructions in Section 2 to access the SOEP data and to place the source files of each version in the appropriate subdirectory.
2. Run `Master_Part1_StataDataPrep.do`.
   Either execute the script using the Terminal/System prompt from the root directory of the repository or open the script by double-clicking (or right-click + Open in Stata) and run it from the Stata GUI. This procedure ensures appropriate setting of the working directory. All paths in the repository are defined relative to this root location.
   - Approx. run time = 5 mins.
   - This script sets the directories and calls the following sub programs:
     - `setup.do` sets the directories for the repository and installs all required Stata packages into a Stata environment (`StataCode/01_code/ado/`).
     - `gen_estim_sample.do` generates the main estimation sample and all additional datasets required in auxiliary regressions. Subprogram `gen_baseline_raw.do` retrieves the relevant cross-sections from the SOEP version 33 data and merges them to a panel dataset. `gen_wealth_raw.do` additionally retrieves information on wealth from the SOEP version 27. The resulting datasets are combined and saved in the folder `Data/SOEP_confid/derived_confid/`.

Descriptive statistics referenced in Footnote 8 of the manuscript and in the text of the Web Appendix are written to Sheets 'Chapter_3' and 'WebAppen_AddResults' in the Excel `CollectedResults.xlsx`, respectively.
Note that any intermediary dataset used in the data preparation procedure is temporarily saved to the folder `Data/tempfiles` and erased once no longer required.

- First stage estimation procedures:
  * `health_profiles.do` estimates health transition probabilities and generates 'Figure_3a_HealthGood.pdf' and 'Figure_3b_HealthBad.pdf' [Figures 3 a/b in main text].
  * `jobsep_prob.do` estimates the involuntary separations model and exports parameter estimates to Sheet 'Tab_SWA3_EmplRisk' [Table SWA.3, Panel II in Web Appendix].
  * `estim_longevity.do` estimates heterogeneous survival probabilities and exports associated hazard ratios to Sheet 'Tab_SWA2_Mort' [Table SWA.2 in Web Appendix]. For this purpose the script generates an additional dataset combining information from SOEP version 35 data and lifetables from the HMD.
- All results required for the estimation of the life-cycle model are saved to `MatlabCode/01_input/`.

5.2 Part 2: Estimation of life-cycle model

Notes:

- Computations described in this section were implemented on one node of a high-performance computing cluster (Bennett et al., 2020). The node is equipped with two 16-core Intel Xeon "Skylake" 6130 processors (i.e., 32-cores) and up to 24 GB of RAM per core. The cluster nodes run AlmaLinux 8 and use Slurm (Simple Linux Utility for Resource Management) as the batch scheduler for resource and job management.
- Optimization procedures in the estimation of the life-cycle model include computationally expensive derivations of numerical gradients over 35 parameters. Gradient derivations may be computed in parallel. The available number of workers in the parallel environment is specified in line 49 of control script `Master_Part2_MatlabEstim.m`. Please allocate an efficient amount of workers to optimize runtime of (35+1) computations of the gradient-loop (e.g. 36, 18, 12, ...).
- Computations documented in the manuscript were implemented using Matlab 2021a, 18 cores on one node of the high-performance computing cluster referenced above, and required a minimum of 3 GB RAM per core.
- Alternatively, the estimation procedures can be run on a desktop machine. Required steps for both procedures are detailed below.
- Numerical precision may vary due to differences in hardware configurations, especially when comparing high-performance computing clusters to standard desktop environments. These variations stem from, e.g., differences in processor architectures, floating-point arithmetic implementations, and compiler optimizations across systems. If present, such differences are expected to remain minimal.

Steps - Computing Cluster:

1. Transfer the entire repository to the user's home directory on the cluster. Note that the cluster environment may have different directory structures and file access permissions. If necessary, the required disk space can be reduced by not transferring the directories of the SOEP source files `Data/SOEP_confid/SOEPxx`. Original data source files are no longer required after completion of Block 1.
2. Configure a shell script to run the estimation job. The script should include commands that set the wall-clock time, control the distribution of computational resources and sets the path to the root of the repository. An example script is provided in `RunModelEstim.sh`. Make sure to adjust this or a similar file to your requirements.
3. Submit the estimation job using the terminal command `sbatch RunModelEstim.sh`.
   - Approx. run time = 50 hours.
   - The script `Master_Part2_MatlabEstim.m` calls the estimation procedures of the main life-cycle model. Details on executed procedures and generated results are provided below when describing execution in a desktop environment.

4. After the code has finished running, transfer the updated repository out of the cluster back to the user's local environment.

Steps - Desktop Environment:

1. Make sure the current working directory is set to the root of the repository. Note that Matlab sets the startup folder based on the way the program is started. Behavior may vary depending on the operating system. For details, see `https://de.mathworks.com/help/matlab/matlab_env/matlab-startup-folder.html`.
   It is recommended to run the Master script using the Terminal/System prompt from the project root directory. This ensures that the working directory is set correctly. Alternatively, navigate to the project directory in the Matlab GUI before running the Master script.

2. Run `Master_Part2_MatlabEstim.m`.
   The script sets up the model environment, estimates the life-cycle model, displays and exports the estimation results, and estimates robustness calibrations. Key subprograms are referenced below. Generated outputs are saved to `MatlabCode/02_output/`. Results documented in the manuscript and online appendix are exported to the Excel file `CollectedResults.xlsx`.

   (a) Setup of model environment:
      - `calibrate.m` initializes the model environment and sets the calibrated parameters.
      - `dataprep.m` loads all required data inputs into the model environment and generates 'Figure_3c_MortalityRisk.svg' [Figure 3(c) in main text].
      - `vargen.m` generates variable and choice-set arrays.

   (b) Estimation of the life-cycle model via Maximum Likelihood is executed as optimization over two functions. First, the earnings process is estimated separately to obtain sensible starting values for the main estimation using `wageloglik.m`. Second, the life-cycle model is estimated using `totloglik.m`. Point estimates of all structural parameters and associated standard errors are saved to file `estim_params.mat`.

   (c) Display and export of estimation results: `paramdisp.m` displays the estimation results in the command window and writes all values referenced in manuscript to Excel file `CollectedResults.xlsx`.
      - Parameters of the utility function, wage equation, education model and type probabilities are written to Sheet 'Tab_1_KeyParams' [Table 1 in main text].
      - Predicted job offer and involuntary job separation probabilities (incl. standard errors) are written to Sheet 'Tab_3_OffSep' [Table 3 in main text].
      - Parameter estimates for the job offer risk model are exported to Sheet 'Tab_SWA3_EmplRisk' [Table SWA.3, Panel I in Web Appendix].

   (d) Robustness checks – alternative calibration of discount factor and risk aversion parameters: re-estimation of the life-cycle model follows the same procedure as described above. Resulting parameters estimates are saved to files `estim_params_bXX_gXX.mat`, where suffixes indicate the underlying calibration of the discount factor β and CRRA parameter γ.

   Note that procedures described in sections (b)-(d) can be run independently of each other given the order of execution is preserved. Switches to toggle executions can be set in lines 19-32 of the Master script.

5.3 Part 3: Simulation exercises

Notes:

- Computations described in this section were implemented using Matlab 2024a on a `MacBook Pro M1 2021 16GB` machine.

- Simulation procedures can be executed in a parallel environment. Each scenario includes the simulation of a sample of R = 50,000 life-cycles. Computation of individual life-cycles can be distributed to multiple workers given the full sample is split evenly into Z subsamples. E.g., computations described in this section used 8 cores, allocating the simulation of R/Z = 50,000/8 = 6,250 life-cycles to each worker.

- Size of simulated samples R is set in `calibrate.m`, and both Z and the number of available workers is specified in lines 61-64 of `Master_Part3_MatlabSim.m`.

- All simulated datasets are saved as `.txt` files to directory `Data/SimData/` for use in the analysis described in the next section. Other output files that are only required for the simulation procedures are saved as `.mat` files to directory `MatlabCode/02_output/`. Results presented in the manuscript and online appendix are exported to associated Sheets in Excel book `CollectedResults.xlsx`.

Steps:

1. Set the current working directory to the root of the repository (see previous Section 5.2, steps in desktop environment). It is recommended to run the Master script using the Terminal/System prompt from the project root.

2. Run `Master_Part3_MatlabSim.m`.
   The script sets up the model environment, generates all simulations for counterfactual scenarios and robustness checks, and computes welfare effects for the policy reform scenario. Execution is split into sections described below. In principle, sections can be executed independently of each other given the order of execution is preserved. Switches to toggle executions can be set in lines 25-43 of the Master script.

   - Approx. run time = 3 hours.

   (a) Setup of model environment (see Section 5.2).

   (b) Static microsimulation using the estimation sample: `samplus.m` runs a static microsimulation that applies the estimation sample to the specification of the proposed life-cycle model environment. The resulting dataset is saved to `Data/SOEP_confid/derived_confid/esample_plus.txt`.

   (c) Simulation of baseline scenario: `simcontrol_baseline.m` constructs the simulated sample for the baseline scenario (`data_baseline.txt`) using the estimated life-cycle model and generates 'Figure_SWA7_Educ Fit.svg' [Figure SWA.7 in Web Appendix].

   - Note that `drawgen.m` is called before each simulation procedure to generate identical pseudo-random draws of stochastic model elements across scenarios.

   (d) Simulation of involuntary separations: `simcontrol_sepsim.m` adds explicit simulation of the stochastic involuntary separations process.

   (e) Simulation of counterfactual risk scenarios: `simcontrol_riskscen.m` constructs simulated samples for the employment and health risk scenarios labeled A-C and saved to files `data_scenario_A/B/C.txt`.

   (f) Simulation of counterfactual tax policy scenario: `simcontrol_taxpol_bisec.m` constructs simulated samples for the policy reform scenarios. Simulated datasets are saved to files `data_scenario_D/E.txt`. A simple bisection root finding procedure is used to calibrate parameters of the lifetime tax function and ensures the reform is revenue neutral. Initial bounds of search procedure are specified in line 109. Solution values for parameters [Table 11 in main text] are exported to Sheet 'Tab_11_PolSim'. Specification and solution paths of search procedures are saved to `TaxPolSol_ScenD/E.mat`.

   (g) Welfare effects of policy reform: `control_taxpol_analysis.m` computes the welfare effects of the tax reform and the share of 'winners'. Results are written to Sheet 'Tab_12_Welfare' [Table 12 in main text].

   (h) Simulations to test robustness to calibration of preference parameters: `simcontrol_robprefparam.m` constructs simulated samples using the estimated model for different calibrations. Resulting datasets are saved to files `data_bXXgXX.txt` following the naming scheme described in Section 5.2.

5.4 Part 4: Goodness of fit and analysis

Notes:

- Computations described in this section were implemented using Stata version 17 (SE-Standard Edition) on a `MacBook Pro M1 2021 16GB` machine.

- All generated figures are saved to directory `StataCode/02_figures/`. Results for tables in the manuscript and online appendix are exported to Excel book `StataCode/03_tables/CollectedResults.xlsx`.

Steps:

1. Run `Master_Part4_StataAnalysis.do`.
   Either execute the script using the Terminal/System prompt from the root directory of the repository or open the script by double-clicking (or right-click + Open in Stata) and run it from the Stata GUI. This procedure ensures appropriate setting of the working directory. All paths in the repository are defined relative to this root location.

   - Approx. run time = 2 hours.
   - All called subprograms and generated outputs are described below.

   (a) Data preparation for analysis: `analysis_dataprep.do`
      - Preparation of estimation sample including the results of static microsimulation for use in the analysis of the goodness of fit of the estimated model.
      - Simulated datasets are prepared for analysis: restrictions as described in the main text are applied; all required employment and income measures are generated (see `proginc.do`).
      - Constructs a subsample of simulated data that matches the age structure observed in the estimation sample for the goodness of fit analysis. This includes random sampling from the full sample of simulated life-cycles in the baseline scenario. For each of the observed individuals in the estimation sample, five randomly selected individuals from the simulated sample who have the same years of education are matched. The number of selected individuals from the simulation sample can be adjusted in lines 354 and 512.
      - Generates permutation matrix used to derive order-robust program contributions to redistribution and insurance effects of the tax-and-transfer system. The matrix is saved to file `Data/SimData/perms.dta`.

   (b) Model and sample descriptives: `sample_descript.do`
      - Generates 'Figure_1a_TTETax.pdf', 'Figure_1b_AvETax.pdf' that illustrate the specification of the tax system in the model [Figures 1(a/b) in main text].
      - Descriptive statistics for the SOEP sample are written to Sheet 'Tab_SWA1_Sample' [Table SWA.1 in Web Appendix].

   (c) In-sample goodness of fit and model validation statistics: `modelfit.do`
      - Generates estimated wage profiles 'Figure_2a_WageProfLow.pdf', Figure_2b_WageProfHigh.pdf' [Figure 2 in main text]. The joint distribution of education and productive ability types is written to Sheet 'Tab_2_Corr' [Table 2 in main text].
      - Creates figures illustrating the goodness of fit of the following observed and predicted age profiles:
        - 'Figure_4a_FitEmpl.pdf', 'Figure_4b_EarnQuantProf.pdf', 'Figure_4c_FitWealth.pdf' [Figure 4 in main text]
        - 'Figure_SWA2a_FitUE.pdf', 'Figure_SWA2b_FitRet.pdf' [Figure SWA.2 in Web Appendix]
      - Creates figures illustrating the goodness of fit for the following observed and predicted distributions:
        - 'Figure_SWA1a_FitWages.pdf', 'Figure_SWA1b_FitWagesHE.pdf', 'Figure_SWA1c_FitWagesLE.pdf' [Figure SWA.1 in Web Appendix]
        - 'Figure_SWA4a_FitAAE.pdf', 'Figure_SWA4b_FitAAEHE.pdf', 'Figure_SWA4c_FitAAELE.pdf' [Figure SWA.4 in Web Appendix]
        - 'Figure_SWA6a_ObsWealth.pdf', 'Figure_SWA6b_SimWealth.pdf' [Figure SWA.6 in Web Appendix]
      - Generates further in-sample goodness of fit measures on employment:
        - Persistence of unemployment written to Sheet 'Tab_SWA4_LSPers' [Table SWA.4 in Web Appendix]
        - 'Figure_SWA3a_UESurv.pdf', 'Figure_SWA3b_UESurvHE.pdf', 'Figure_SWA3c_UESurvLE.pdf' [Figure SWA.3 in Web Appendix]

  - Shares of involuntary separations written to Sheet 'Tab_SWA6_Invol' [Table SWA.6 in Web Appendix]

- Generates goodness of fit measures on earnings mobility:
  - Rank correlations between labor earnings over time written to Sheet 'Tab_4_RankCorr' [Table 4 in main text]
  - Subprogram `earn_mobil.do` writes labor earnings transition matrices to Sheet 'Tab_SWA5_EarnMob' [Table SWA.5 in Web Appendix]

- Exports validation statistics to Sheet 'Tab_5_Valid' [Table 5 in main text]

(d) Analysis of simulated data – Part 1: `analysis_part1.do`
This script runs the main inequality decomposition analysis and derives insurance and redistributive effects of the tax-transfer system. Called subprograms and generated outputs are detailed below.

- Programs `analysis_part1_decomp.do` and `analysis_part1_decomp_varinc.do` run the decomposition and obtain marginal contributions of programs in all simulated scenarios.

- Results are written to Sheets of Excel `CollectedResults.xlsx` labeled in accordance with the name of the associated simulated dataset and the used measure of inequality.
  - 'Base_[...]' indicating the baseline scenario.
  - Counterfactual scenarios are labeled 'A-E_[...]':
    - Scenario A: Increased job separation risk.
    - Scenario B: Decreased job offer rate.
    - Scenario C: Increased risk of bad health shocks.
    - Scenario D: Tax policy reform with behavior consistent with the baseline scenario.
    - Scenario E: Tax policy reform with behavioral adjustments.
  - Inequality measures are labeled following their characterisation as members of the single parameter Generalized Entropy class: Mean logarithmic deviation (ge0), Theil index (ge1) and half the squared coefficient of variation (ge2). Additionally, we compute results based on the variance of the natural logarithm ('[...]_VarInc').
- Various robustness checks are run and written to Sheets of `CollectedResults.xlsx`: exclusion of interest income ['Base_NoInt_ge1'], correction for negative lifetime earnings ['Base_PosLTEarn_ge1'], working reduced hours ['Base_PT2_ge1', 'Base_PT3_ge1'], alternative calibration of preference parameters ['Base_bXXgXX_ge1'].

- The excel file `CollectedResults.xlsx` includes separate Sheets that organize the numbers as they appear in Tables in the paper:
  - Tables 6&7 in the main text summarize the results of the main decomposition analysis for the baseline scenario [Sheet 'Tab_6_7_MainDecomp'].
  - Table 11 summarises insurance and redistribution effects in the counterfactual tax policy scenarios (labeled D and E) [Sheet 'Tab_11_PolSim'].
  - Table SWA.8 (Web Appendix) shows robustness to excluding capital income [Sheet 'Tab_SWA8_Rob NoInt']
  - Table SWA.9 (Web Appendix) shows robustness of results in Tables 6&7 to alternative measures of inequality [Sheet 'Tab_SWA9_RobIneqMeasures']
  - Tables SWA.10&11 (Web Appendix) show robustness of results in Tables 6&7 to calibration of preference parameters [Sheet 'Tab_SWA10_11_RobPref']

- The sample share of simulated individuals with negative lifetime earnings referenced in Footnote 20 and any additional values reported in the main text of Chapter 5 are written to Sheet 'Chapter5_AddResults'.

- The effect of the tax-and-transfer system on inequality of annual income reported in Footnote 23 is written to Sheet 'Tab_6_7_MainDecomp'.

- Figures used in the graphical analysis of Chapter 5 are generated:
  - 'Figure_5a_InsTax_TaxShare.pdf' and 'Figure_5b_InsTax_YearsWorked.pdf' [Figure 5 in main text]
  - 'Figure_6a_RedTax_TaxShare.pdf', 'Figure_6b_RedTax_Empl.pdf', 'Figure_6c_RedTax_Avg Earn.pdf', 'Figure_6d_RedTax_TaxShare.pdf' [Figure 6 in main text]
  - 'Figure_7a_DBsHealth.pdf', 'Figure_7b_DBsReceipt.pdf' [Figure 7 in main text]
  - 'Figure_8a_SAIns_IncGap.pdf', 'Figure_8b_SAIns_Wealth.pdf' [Figure 8 in main text]
  - 'Figure_9a_SARed_IncGap.pdf', 'Figure_9b_SARed_Wealth.pdf' [Figure 9 in main text]

- Additional figures presented in the supplementary analysis of Web Appendix V are generated:
  - 'Figure_SWA8a_InsTax_AvgEarnPerYear.pdf' and 'Figure_SWA8b_InsTax_SDAvgEarn.pdf' [Figure SWA.8]
  - 'Figure_SWA9a_InsTax_AvgEarnPerYear_NoWS.pdf' and 'Figure_SWA9b_InsTax_SDAvgEarn_NoWS.pdf' [Figure SWA.9]

(e) Analysis of simulated data – Part 2: `analysis_part2.do`
This script runs the analysis of counterfactual risk scenarios presented in Chapter 6. It derives behavioral implications and changes of the inequality mitigating effects of the tax and transfer system as a result of variations in employment or health risk. All generated outputs are detailed below.

- Results for behavioral effects and inequality decompositions for all scenarios are written Sheets of Excel `CollectedResults.xlsx` labeled 'Part2_[...]', where the suffix indicates the used measure of inequality (see point d) above).

- In addition, the Excel book includes separate Sheets that organize the numbers as they appear in Tables in the manuscript:
  - Behavioral responses in the elevated risk environments are summarized in Sheet 'Tab_8_RiskBeh' [Table 8 in main text]
  - Insurance effects of the tax-transfer system are summarized in Sheet 'Tab_9_10_RiskInsures' [Table 9 & 10 in main text]
  - Behavioral effects of the lifetime tax reform: Sheet 'Tab_SWA7_PolSimBeh' [Table SWA.7 in Web Appendix]

- Labor supply effects of the lifetime tax reform are shown in 'Figure_SWA10_PolTaxEmplRate.pdf' [Figure SWA.10 in Web Appendix]

(f) Robustness of results to full-time assumption: `rob_parttime.do` generates figures illustrating the distributional implications and reports robustness of decomposition results.

- 'Figure_SWA5a-f_[...]' show observed and predicted persistence in labor earnings [Figure SWA.5 in Web Appendix]
- Decomposition results are written to Sheet 'Tab_SWA16_RobPT' [Table SWA.16 in Web Appendix]

6 List of tables and programs

Table 1: Mapping of programs to tables, including data source(s)

Table #	Program	Line number	Data source(s)
Table 1	paramdisp.m	171, 178, 183, 187	SOEP27/33/35, HMD
Table 2	modelfit.do	166-173, 199	SOEP27/33/35, HMD
Table 3	paramdisp.m	203, 209, 215, 221	SOEP27/33/35, HMD
Table 4	modelfit.do	1127	SOEP27/33/35, HMD
Table 5	modelfit.do	1171-1173	SOEP27/33/35, HMD
Table 6	analysis_part1_decomp.do	82	SOEP27/33/35, HMD
Table 7	analysis_part1_decomp.do	162-164	SOEP27/33/35, HMD
Table 8	analysis_part2.do	84	SOEP27/33/35, HMD
Table 9	analysis_part2.do	166	SOEP27/33/35, HMD
Table 10	analysis_part2.do	169	SOEP27/33/35, HMD
Table 11	analysis_part1_decomp.do	82, 162-164	SOEP27/33/35, HMD
	analysis_part2.do	84	SOEP27/33/35, HMD
	simcontrol_taxpol_bisec.m	336, 360	SOEP27/33/35, HMD
Table 12	control_taxpol_analysis.m	62, 64, 129, 131	SOEP27/33/35, HMD
Table SWA.1	sample_descript.do	144-147	SOEP27/33
Table SWA.2	estim_longevity.do	661-667	SOEP35, HMD
Table SWA.3 (I)	paramdisp.m	194	SOEP27/33/35, HMD
Table SWA.3 (II)	jobsep_prob.do	119-125	SOEP33
Table SWA.4	modelfit.do	822-858	SOEP27/33/35, HMD
Table SWA.5	earn_mobil.do	144, 283	SOEP27/33/35, HMD
Table SWA.6	modelfit.do	1298-1313	SOEP27/33/35, HMD
Table SWA.7	analysis_part2.do	84	SOEP27/33/35, HMD
Table SWA.8	analysis_part1_decomp.do	82, 162-164	SOEP27/33/35, HMD
Table SWA.9	analysis_part1_decomp.do	82, 162-164	SOEP27/33/35, HMD
	analysis_part1_decomp_varinc.do	96, 214-216	SOEP27/33/35, HMD
Table SWA.10	analysis_part1_decomp.do	82, 162-164	SOEP27/33/35, HMD
Table SWA.11	analysis_part1_decomp.do	82, 162-164	SOEP27/33/35, HMD
Table SWA.12	analysis_part2.do	166, 169, 349, 352	SOEP27/33/35, HMD
Table SWA.13	analysis_part1_decomp.do	82, 162-164	SOEP27/33/35, HMD
Table SWA.14	analysis_part1_decomp.do	82, 162-164	SOEP27/33/35, HMD
Table SWA.15	analysis_part1_decomp_varinc.do	96, 214-216	SOEP27/33/35, HMD
Table SWA.16	rob_parttime.do	279	SOEP27/33/35, HMD

Table 2: Mapping of programs to figures, including data source(s)

Figure #	Program	Line number	Data source(s)
Figure 1 (a)	sample_descript.do	88	None (policy rules)
Figure 1 (b)	sample_descript.do	99	None (policy rules)
Figure 2 (a)	modelfit.do	116	SOEP27/33/35, HMD
Figure 2 (b)	modelfit.do	103	SOEP27/33/35, HMD
Figure 3 (a)	health_profiles.do	101	SOEP33
Figure 3 (b)	health_profiles.do	112	SOEP33
Figure 3 (c)	dataprep.m	111	SOEP35, HMD
Figure 4 (a)	modelfit.do	266	SOEP27/33/35, HMD
Figure 4 (b)	modelfit.do	1069	SOEP27/33/35, HMD
Figure 4 (c)	modelfit.do	278	SOEP27/33/35, HMD
Figure 5 (a)	analysis_part1.do	236	SOEP27/33/35, HMD
Figure 5 (b)	analysis_part1.do	256	SOEP27/33/35, HMD
Figure 6 (a)	analysis_part1.do	465	SOEP27/33/35, HMD
Figure 6 (b)	analysis_part1.do	486	SOEP27/33/35, HMD
Figure 6 (c)	analysis_part1.do	503	SOEP27/33/35, HMD
Figure 6 (d)	analysis_part1.do	520	SOEP27/33/35, HMD
Figure 7 (a)	analysis_part1.do	600	SOEP27/33/35, HMD
Figure 7 (b)	analysis_part1.do	617	SOEP27/33/35, HMD
Figure 8 (a)	analysis_part1.do	706	SOEP27/33/35, HMD
Figure 8 (b)	analysis_part1.do	728	SOEP27/33/35, HMD
Figure 9 (a)	analysis_part1.do	750	SOEP27/33/35, HMD
Figure 9 (b)	analysis_part1.do	776	SOEP27/33/35, HMD
Figure SWA.1 (a)	modelfit.do	352	SOEP27/33/35, HMD
Figure SWA.1 (b)	modelfit.do	365	SOEP27/33/35, HMD
Figure SWA.1 (c)	modelfit.do	378	SOEP27/33/35, HMD
Figure SWA.2 (a)	modelfit.do	290	SOEP27/33/35, HMD
Figure SWA.2 (b)	modelfit.do	302	SOEP27/33/35, HMD
Figure SWA.3 (a)	modelfit.do	772	SOEP27/33/35, HMD
Figure SWA.3 (b)	modelfit.do	796	SOEP27/33/35, HMD
Figure SWA.3 (c)	modelfit.do	784	SOEP27/33/35, HMD
Figure SWA.4 (a)	modelfit.do	397	SOEP27/33/35, HMD
Figure SWA.4 (b)	modelfit.do	411	SOEP27/33/35, HMD
Figure SWA.4 (c)	modelfit.do	425	SOEP27/33/35, HMD
Figure SWA.5 (a)	rob_parttime.do	126	SOEP27/33/35, HMD
Figure SWA.5 (b)	rob_parttime.do	138	SOEP27/33/35, HMD
Figure SWA.5 (c)	rob_parttime.do	150	SOEP27/33/35, HMD
Figure SWA.5 (d)	rob_parttime.do	164	SOEP27/33/35, HMD
Figure SWA.5 (e)	rob_parttime.do	176	SOEP27/33/35, HMD
Figure SWA.5 (f)	rob_parttime.do	188	SOEP27/33/35, HMD
Figure SWA.6 (a)	modelfit.do	1338	SOEP27/33/35, HMD
Figure SWA.6 (b)	modelfit.do	1349	SOEP27/33/35, HMD
Figure SWA.7	simcontrol_baseline.m	64	SOEP27/33/35, HMD
Figure SWA.8 (a)	analysis_part1.do	320	SOEP27/33/35, HMD
Figure SWA.8 (b)	analysis_part1.do	338	SOEP27/33/35, HMD
Figure SWA.9 (a)	analysis_part1.do	358	SOEP27/33/35, HMD
Figure SWA.9 (b)	analysis_part1.do	377	SOEP27/33/35, HMD
Figure SWA.10	analysis_part2.do	450	SOEP27/33/35, HMD

References

Bennett, L., Melchers, B., and Proppe, B. (2020). Curta: A general-purpose high-performance computer at ZEDAT, Freie Universität Berlin

Goebel, J., Grabka, M.M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2019). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik*, 239(2):345–360

HMD (2024). Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at www.mortality.org (data downloaded on April 15, 2024)

SOEP (2011). Socio-Economic Panel (SOEP), data for years 1984-2010, version 27. doi:10.5684/soep.v27

SOEP (2017). Socio-Economic Panel (SOEP), data for years 1984-2016, version 33. doi:10.5684/soep.v33

SOEP (2019). Socio-Economic Panel (SOEP), data for years 1984-2018, version 35. doi:10.5684/soep.v35

Wagner, G., Frick, J., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) - Scope, evolution and enhancements. *Schmollers Jahrbuch*, 127(1):139–169