Replication for Marriage Market and Labor Market Sorting

Paula Calvo, Ilse Lindenlaub, Ana Reynoso

July 2023

OVERVIEW

Software Requirements

- Stata Versions: 15.0 and 17.0.
    - Additional packages used: binscatter, coefplot, egenmore, estadd, estpost, eststo, esttab, estout, ivreg2, xtivreg2, outreg2, outtable, unique, ftools, reghdfe.
- MATLAB Version : 2021b
- Memory and Runtime Requirements: The code was last run on July 21, 2023. All do-files in Stata were run on a personal laptop—Windows or MAC—with 32G memory (each code runs in less than a minute); the structural estimation and post-estimation exercises were run in MATLAB on a personal computer with 16G memory (Windows) with a runtime of less than 40 hours.

Data Sources and Availability

We use data from three sources:

1. German Socio Economic Panel (GSOEP) from the *German Institute for Economic Research* (DIW). Due to data protection regulations, the data cannot be made publicly available. However, the data are available for scientific research free of charge and any researcher can apply for and obtain access by signing a data distribution contract. The data distribution contract can be requested with a form, available at: Link. For further information, contact the SOEPhotline at +49-30-89789-292 or at soepmail@diw.de.

   Note: Due to these restrictions, the working datasets and associated codes that allow to replicate all the results in this artiche based on the GSOEP are hosted at the SOEP Reanalysis Archive at DIW Berlin. These datasets were created based on v35 of the GSOEP raw data, doi:10.5684/soep.v35. For instructions on how to access our working data and codes at the SOEP Reanalysis Archive, see contact information of the SOEP department at DIW Berlin here.

2. BIBB/BAuA Employment Survey of the Working Population on Qualification and Working Conditions in Germany 2012 (*Federal Institute for Vocational Education and Training*); access to the micro-data with 4-digit occupations is not public. For information on how to obtain access, see this Link.

3. German Time Use Survey (GTUS) from the *German Federal Statistical Office*; public use files can be downloaded from this Link.

Overview of the Replication Directory

Our project folder `Replication` has the following sub-folders:

1. `Empirics` : Here we conduct the empirical analysis of the paper, which—due to data protection issues—needs to be replicated at the SOEP Reanalysis Archive at DIW Berlin (see above) upon applying for the data.
   a. `Data` : Contains the datasets used to produce the empirical output of the paper.
   b. `Dofiles` : Contains the do-files to generate the empirical output of the paper.
   c. `Output` : Stores the output of the do-files.

2. `Estimation` : Runs the estimation and post-estimation analysis
   a. `Model` : Runs the estimation and post-estimation. Sub-folders:

      - `Baseline` : Contains the files to reproduce estimation of the model in the period 2010-2016.
      - `Past` : Contains the files to reproduce estimation of the model in the period 1990-1996.
      - `Post_estimation` : Contains the following sub-folders with files to compile estimation results and perform the quantitative analysis of the paper:

        - `SE and Sensitivity` : Produces standard errors of estimates for estimation period 1990-1996 and 2010-2016.
        - `Fit and Validation` : Creates tables and figures with estimation results and model validation.
        - `Over_time` : Produces inequality changes over time and decomposition into drivers.
        - `Elasticities` : Creates elasticities of wage inequality with respect to marriage and labor market sorting.
        - `Comparative Statics` : Solves our model at various values of parameters.
        - `Uniqueness` : Plots the deviations from simulated wage functions and our equilibrium wage function for different starting points of the fixed point computation.
    - `Heterogeneous_Home_Production` : contains programs to re-estimate the model for different types of households (for sample with and without children; and for the specification with heterogeneous home production parameters—both for period 1990-1996 and 2010-2016).

b. `Dofiles` : Contains the do-files to generate the data inputs used in model estimation and post-estimation exercises. The results obtained from these codes can be replicated at the SOEP Reanalysis Archive at DIW Berlin (see above).

c. `Data` : Contains data used as intermediate inputs in the model estimation. Sub-folders:

- `moments` : Stores the data moments, bootstrapped moments, and confidence intervals of moments.
- `validation` : Stores untargeted moments and auxiliary datasets used to produce validation figures.
- `data_inputs_for_estimation` : Uses as inputs the data files in `moments` to produce the matlab files containing the data inputs, used in the estimation. These files are stored in sub-folder `data_inputs_for_estimation/Output` .

3. `Readme` : Contains this Readme file.

Overview of Table and Figure Replication

Here we list the source code for each table and figure in the paper and online appendix. Each file is described in detail below.

Tables

Number	Title	Data/Source	Code	Detailed Instructions
1	Marriage Matching Frequencies by Education	GSOEP	`empirics_CLR_1.do`	Section B
2	Estimated Parameters	Estimation	`post_estimation_CLR_3.m`	Section C
3	Untargeted Moments: Marriage Matching Frequencies—Model and (Data)	Estimation, GSOEP	`post_estimation_CLR_3.m`	Section C
4	Gender and Household Inequality	Estimation, GSOEP	`post_estimation_CLR_3.m`	Section C
5	Elasticity of Inequality with Respect to Sorting	Estimation	`post_estimation_CLR_5.m`	Section C
A.1	Calibrated Parameters	Estimation	`post_estimation_CLR_3.m`	Section C
A.2	Targeted Moments	Estimation, GSOEP	`post_estimation_CLR_3.m`	Section C
A.3	Estimated Parameters: 1990-1996 versus 2010-2016	Estimation, GSOEP	`post_estimation_CLR_3.m`	Section C
A.4	Data and Model Moments: 1990-1996 versus 2010-2016	Estimation, GSOEP	`post_estimation_CLR_3.m`	Section C
O.1	Marriage Matching Frequencies by Education (1990-1996)	GSOEP	`empirics_CLR_1.do`	Section B
O.2	Marital Sorting Parameters (2010-2016)	GSOEP	`empirics_CLR_1.do`	Section B
O.3	Marital Sorting Parameters (1990-1996)	GSOEP	`empirics_CLR_1.do`	Section B
O.4	Labor Market Sorting and Marriage Market Sorting	GSOEP	`empirics_CLR_1.do`	Section B
O.5	Labor Market Sorting and Marriage Market Sorting, by Education Level	GSOEP	`empirics_CLR_1.do`	Section B
O.6	Complementarity in Home Production Hours	GSOEP	`empirics_CLR_2.do`	Section B
O.7	Complementarity in Labor Market Hours	GSOEP	`empirics_CLR_2.do`	Section B
O.8	Complementarity in Home Production Hours, by Male	GSOEP	`empirics_CLR_2.do`	Section B

Number	Title	Data/Source	Code	Detailed Instructions
	Partner's Education Level			
O.9	Complementarity in Labor Market Hours, by Male Partner's Education Level (IV Results)	GSOEP	`empirics_CLR_2.do`	Section B
O.10	Moments (illustrative table)	—	—	
O.11	Selection Regression	GSOEP	`empirics_CLR_3.do`	Section B
O.12	Wage Regression	GSOEP	`empirics_CLR_3.do`	Section B
O.13	Worker Distribution of s-Types by Education	GSOEP	`empirics_CLR_1.do`	Section B
O.14	Decomposing Changes in the Aggregate Home Production Correlation (Tasks)	GTUS	`empirics_CLR_4.do`	Section B
O.15	Decomposing Changes in Home Production Correlation: Couples with and without Children	GSOEP	`empirics_CLR_1.do`	Section B
O.16	Decomposing Changes in Home Production Correlation: Couples of Different Education	GSOEP	`empirics_CLR_1.do`	Section B
O.17	Estimated Parameters by Presence of Children: 1990-1996 versus 2010-2016	Estimation	`hhp_CLR_1.m`	Section C
O.18	Estimated Parameters by Couples' Skill Type: 1990-1996 versus 2010-2016	Estimation	`hhp_CLR_1.m`	Section C
O.19	Complementarity in Home Production Hours: Childcare vs. Housework	GSOEP	`empirics_CLR_2.do`	Section B

Figures

Number	Title	Data/Source	Code	Detailed Instructions
1	Labor Market Matching Function (left) and Labor Market and Marriage Market Sorting (right)	GSOEP, BIBB	`empirics_CLR_1.do`	Section B
2	Time Allocation and Marriage Sorting	GSOEP	`empirics_CLR_1.do`	Section B
3	The Decision Stages (illustrative figure)	—	—	
4	Model Fit: Model Moments (red) with Data Confidence Intervals (blue)	Estimation	`post_estimation_CLR_3.m`	Section C
5	Labor Market Matching Function, Original (left) and with Hours Partialled Out (right)	Estimation, GSOEP, BIBB	`post_estimation_CLR_3.m`	Section C
6	Labor Market Sorting and Marriage Market Sorting (left); Home Production Hours and Marriage Market Sorting (right)	Estimation, GSOEP, BIBB	`post_estimation_CLR_3.m`	Section C
7	Inequality Changes over Time (left); Sorting and Hours Changes over Time (right).	Estimation, GSOEP	`post_estimation_CLR_4.m`	Section C
8	Mechanism behind Inequality Changes	Estimation	`post_estimation_CLR_4.m`	Section C
9	Contribution of Detailed Home Production Tasks to Aggregate Home Production Correlation, Cross-section of 2012/13 (left) and over Time between 1991/92 and 2012/13 (right)	GTUS	`empirics_CLR_4.do`	Section B
A.1	Inequality Changes Over Time: Detailed Decomposition	Estimation	`post_estimation_CLR_4.m`	Section C
O.1	Labor and Marriage Market Sorting, by Education Level	GSOEP	`empirics_CLR_1.do`	Section B
O.2	Complementarities in Home Production Hours, by Male Partner's Education Level	GSOEP	`empirics_CLR_1.do`	Section B
O.3	Complementarities in Labor Market Hours, by Male Partner's Education Level	GSOEP	`empirics_CLR_1.do`	Section B
O.4	Weekly Mean Hours by Gender, Education and Marital Status	GSOEP	`empirics_CLR_1.do`	Section B
O.5	Part-Time Wage Penalties Relative to Men who Work Full-Time (left); Share of Workers in Full-/Part-Time Work (right)	GSOEP	`empirics_CLR_1.do`	Section B

Number	Title	Data/Source	Code	Detailed Instructions
O.6	Difference Between Our Equilibrium Wage Function and the Convergent Wage Function Resulting from 60 Alternative Initial Guesses	Model	`post_estimation_CLR_7.m`	Section C
O.7	Principal Component Analysis	GSOEP	`empirics_CLR_3.do`	Section B
O.8	Comparative Statics (ρ)	Model	`post_estimation_CLR_6.m`	Section C
O.9	Comparative Statics (θ)	Model	`post_estimation_CLR_6.m`	Section C
O.10	Comparative Statics (ψ)	Model	`post_estimation_CLR_6.m`	Section C
O.11	Home Production over Time: Correlation of Spouses' Hours (left); Task Shares (right)	GTUS	`empirics_CLR_4.do`	Section B
O.12	Correlation of Spouses' Hours by Couple Type: Children (left); Education (right)	GSOEP	`empirics_CLR_1.do`	Section B
O.13	Contribution of Different Education Groups to the Within-Component of the Aggregate Home Production Correlation, Cross-Section of 2010/16 (left) and Over Time 1990/96-2010/16 (right)	GSOEP	`empirics_CLR_1.do`	Section B
O.14	Model Fit: Model Moments (red) with Data Confidence Intervals (blue)	Estimation	`hhp_CLR_1.m`	Section C
O.15	Model Fit: Model Moments (red) with Data Confidence Intervals (blue)	Estimation	`hhp_CLR_1.m`	Section C

DETAILED REPLICATION INSTRUCTIONS

A. Data Preparation Instructions

Folder: `Empirics/Data`

- Our main dataset for the empirical analysis is `SOEP_dataset`, which can be accessed via the SOEP Reanalysis Archive (please see above). It was created based on the raw data of the GSOEP (wave SOEP-Core v35 doi:10.5684/soep.v35), which is not publicly available. From this wave of the SOEP, we used the following datasets: `ppathl` (individual's demographic information), `pgen` (labor market information), `hpathl`, `hgen`, `hbrutto` (household information), `pbrutto` (relationship between households members), `pl` (wage information, birth of children, and hours in different activities), `biobirth` (birth histories), and `biomarsy` (marital histories). In Online Appendix OD.2, we provide details regarding sample construction and restrictions as well as variable definitions; see also Online Appendix OE.3 and Online Appendix OE.4.
  Additionally, we merged into this data the share of children covered by childcare slots by state and year, based on data from the Federal Statistics Office of Germany (Statistisches Bundesamt; data can be obtained here). To construct our variable of share of children covered by childcare slots, we combined data from Tables 22541-0002 (Children in daycare centers) and 12411-0012 (Population), between 2006 and 2016.
- We use time use data from the GTUS (waves 1991/92 and 2012/13). Here we included the raw data in the replication package because it is publicly available: `zbe_ps`, `zbe_sum`, `zve13_puf_pers`, `zve13_puf_sum`.
- We use data on occupational task requirements to construct occupational types based on the BIBB survey (wave 2012), saved in `BIBB_dataset`. These data, based on raw data on occupational tasks at the 4-digit occupation level (data file `ZA5657_v4-0-0`), are not publicly available. Therefore, our code to construct occupational tasks at the 4-digit level was run by the BIBB Institute. We then used these data to construct `BIBB_dataset` (see Online Appendix OE.4). Dataset `BIBB_dataset` can be accessed via the SOEP Reanalysis Archive (please see above).

Folder: `Estimation/Data`

- Using data from the same GSOEP wave and raw data files described above, we also constructed our dataset for estimation `SOEP_estimation_household`. A detailed description of how we create this estimation sample and the relevant variables can be found in Online Appendix OD.3. This dataset contains the relevant variables to create the data moments used in the estimation. Similarly, we constructed the dataset `SOEP_estimation_individual`, which contains the same sample as `SOEP_estimation_household`, but at the individual rather than at the household level. Both `SOEP_estimation_household` and `SOEP_estimation_individual` can be accessed via the SOEP Reanalysis Archive.

B. Instructions for Replicating the Empirical Analysis

Main Folder

`Empirics`

Description

To reproduce all Tables and Figures from the empirical analysis of the paper (included in Section 2, Section 6, Online Appendix OA, OE, OG, and OH) the following do-files have to be run:

1. `CLR_empirics_1.do`
2. `CLR_empirics_2.do`
3. `CLR_empirics_3.do`
4. `CLR_empirics_4.do`

Note: Due to data restrictions, the do-files `CLR_empirics_1.do`, `CLR_empirics_2.do`, and `CLR_empirics_3.do` can only be run accessing the SOEP Reanalysis Archive at DIW Berlin. Please see above for instructions on how to access this archive and our data.

List of Programs and Produced Output

Folder: `Empirics/Dofiles`

File: `empirics_CLR_1.do`

- **Description.** Produces most of the empirical analysis of the paper (facts, distributions, and decompositions) that are based on GSOEP data.
- **Inputs.** `SOEP_dataset.dta`, loaded from folder `Empirics/Data`; and `SOEP_estimation_individual.dta`, loaded from folder `Estimation/Data`.
- **How to run:** To run this do-file, update the global `path_of_replicator` at the beginning of file `empirics_CLR_1.do`, using the working directory of the replicator.
- **Outputs.** TABLES: 1, O.1., O.2., O.3., O.4, O.5, O.13, O.15, O.16. FIGURES: 1, 2, O.1, O.2., O.3, O.4, O.5, O.12, O.13.; saved to folder `Empirics/Output`.

File: `empirics_CLR_2.do`

- **Description.** Produces empirical results associated with the correlation between partners' hours, using regression analysis.
- **Inputs.** `SOEP_dataset.dta`; loaded from folder `Empirics/Data`.
- **How to run:** To run this do-file, update the global `path_of_replicator` at the beginning of file `empirics_CLR_2.do`, using the working directory of the replicator.
- **Outputs.** TABLES: O.6, O.7, O.8, O.9, O.19; saved to folder `Empirics/Output`.

File: `empirics_CLR_3.do`

- **Description.** Runs the selection regression and the panel wage regression; this code also creates the occupational types. See Appendix OE.3 and OE.4 of the paper for details.
- **How to run:** To run this do-file, update the global `path_of_replicator` at the beginning of file `empirics_CLR_3.do`, using the working directory of the replicator.
- **Inputs.** `SOEP_dataset` and `BIBB_dataset`; loaded from folder `Empirics/Data`.
- **Outputs.** TABLES: O.11, O.12, FIGURES: O.7; saved to folder `Empirics/Output`.

File: `empirics_CLR_4.do`

- **Description.** Conducts the time use analysis (Section 6 of the paper and Online Appendix OG) based on the GTUS.
- **How to run:** To run this do-file, update the global `path_of_replicator` at the beginning of file `empirics_CLR_4.do`, using the working directory of the replicator.
- **Inputs.** `zbe_ps.dta`, `zbe_sum.dta`, `zve13_puf_sum.dta`, `zve13_puf_pers.dta`; loaded from folder `Empirics/Data`.
- **Outputs.** FIGURES: 9, O.11; TABLES: O.14; saved to folder `Empirics/Output`.

C. Instructions for Replicating the Structural Estimation

Main Folder

`Estimation/Model`

Description

To produce all estimation output and post-estimation results (Sections 4, 5, 6 of the paper; and Appendix D and E; and Online Appendix OF and OG), run the following codes in this order:

1. `Baseline/estimation_CLR_1.m`
2. `Past/estimation_CLR_2.m`
3. `Post_estimation/SE and Sensitivity/post_estimation_CLR_1.m` and `Post_estimation/SE and Sensitivity` / `post_estimation_CLR_2.m`
4. `Post_estimation/Fit and validation/post_estimation_CLR_3.m`
5. `Post_estimation/Over_time/post_estimation_CLR_4.m`
6. `Post_Estimation/Elasticities/post_estimation_CLR_5.m`
7. `Post_Estimation/Comparative Statics/post_estimation_CLR_6.m`
8. `Post_Estimation/Uniqueness/post_estimation_CLR_7.m`

Note: The entire estimation and post-estimation analysis in Section C. of the Readme file can be reproduced from the replicator's personal computer (no need to do this in the SOEP Reanalysis Archive). Also, this analysis can be reproduced without constructing the estimation inputs from scratch. If there is interest in reconstructing the estimation inputs, see Section D. below.

List of Programs and Produced Output

Folder `Estimation/Model/Baseline`

File: `estimation_CLR_1.m`

- **Description.** This code runs the our baseline model estimation (period 2010-2016).
- **Inputs.** `Baseline/Functions/inputs_present.mat` (created in `inputs.m` in `Estimation/Data/data_inputs_for_estimation` ; see Section D. of the Readme file for details.)
- **Auxiliary codes.** All functions stored in `Baseline/Functions` .
- **How to run.** Update the `replication_location` at the beginning of file `estimation_CLR_1.m` using the working directory of the replicator, and run this file.
- **Outputs.** `outputs_present_restud.mat` ; stored in `Model/Baseline/Output`

Folder `Estimation/Model/Past`

File: `estimation_CLR_2.m`

- **Description.** This code runs our model estimation for period 1990-1996.
- **Inputs.** `Past/Functions/inputs_past.mat` (created in `inputs.m` in `Estimation/Data/data_inputs_for_estimation` ; see Section D. of the Readme file for details.)
- **Auxiliary codes.** All functions stored in `Past/Functions` .
- **How to run.** Update the `replication_location` at the beginning of file `estimation_CLR_2.m` using the working directory of the replicator, and run this file.
- **Outputs.** `outputs_past_restud.mat` ; stored in `Model/Past/Output`

Folder `Estimation/Model/Post_estimation/SE and Sensitivity`

Files: `post_estimation_CLR_1.m` and `post_estimation_CLR_2.m`

- **Description.** These codes produce the standard errors of our model estimates for the baseline (code: `post_estimation_CLR_1.m`) and the past period (code: `post_estimation_CLR_2.m`).
- **Inputs.** `Baseline/Outputs/outputs_baseline_restud.mat` and `Past/Outputs/outputs_past_restud.mat` and data inputs stored in sub-folder `Functions` (created in `inputs.m` in `Estimation/Data/data_inputs_for_estimation` ; see Section D. of the Readme file for details.)
- **How to run.** Update the `replication_location` at the beginning of files `post_estimation_CLR_1.m` and `post_estimation_CLR_2.m` using the working directory of the replicator, and run these files.
- **Output.s** `SE_pr_restud` and `SE_pa_restud` ; stored in `Post_estimation/SE and Sensitivity/Output` .

Folder `Estimation/Model/Post_estimation/Fit and validation`

File: `post_estimation_CLR_3.m`

- **Description.** This code creates tables and figures of estimation and validation results.
- **Inputs.** Moments and confidence intervals from the data and the following estimation outputs: `outputs_present_restud.mat`, `outputs_past_restud.mat`, `SE_pr_restud.mat`, `SE_pa_restud.mat` (see above).
- **How to run.** Update the `replication_location` path at the beginning of the file `post_estimation_CLR_3.m` using the working directory of the replicator, and run this file.
- **Outputs.** TABLES: 2, 3, 4, A.1, A.2, A.3, A.4; FIGURES: 4, 5, 6.; stored in `Post_estimation/Fit and validation/Output`.

Folder `Estimation/Model/Post_estimation/Over_time`

File: `post_estimation_CLR_4.m`

- **Description.** This code creates the outcomes of our analysis of inequality over time and its decomposition in primitive drivers.
- **Inputs.** Estimation outputs from baseline and past periods (`outputs_present_restud.mat` and `outputs_past_restud.mat`) and moments from the data (stored in `Estimation/Data`).
- **Auxiliary codes.** Uses the functions stored in sub-folder `Functions`.
- **How to run.** Update the `replication_location` at the beginning of the file `post_estimation_CLR_4.m` using the working directory of the replicator, and run this file.
- **Outputs.** FIGURES: 7, 8, A.1.; stored in `Post_estimation/Over_time/Output`.

Folder `Estimation/Model/Post_Estimation/Elasticities`

File: `post_estimation_CLR_5.m`

- **Description.** This code computes the elasticities of inequality with respect to marriage market sorting and labor market sorting.
- **Inputs.** Estimation inputs and outputs from the baseline and past periods (`inputs_present.mat`, `inputs_past.mat`, `outputs_present_restud.mat` and `outputs_past_restud.mat`).
- **Auxiliary codes.** uses the functions stored in `Functions` (and its sub-folders `Functions/LM_PE` and `Functions/MM_PE`).
- **How to run.** Update the `replication_location` at the beginning of the file `Elasticities_CLR_1.m` using the working directory of the replicator, and run this file.
- **Outputs.** TABLE: 5; stored in `Post_estimation/Elasticities/Output`.

Folder `Estimation/Model/Post_Estimation/Comparative Statics`

File: `post_estimation_CLR_6.m`

- **Description.** This code performs numerical comparative statics.
- **Inputs.** Estimation outputs from baseline and past periods (`outputs_present_restud.mat` and `outputs_past_restud.mat`).
- **Auxiliary codes.** Uses the functions stored in sub-folder `Functions`.
- **How to run.** Update the `replication_location` at the beginning of the file `post_estimation_CLR_6.m` using the working directory of the replicator and run this file.
- **Outputs.** TABLES: O.8, O.9, O.10.; stored in `Post_estimation/Comparative Statics/Output`.

Folder `Estimation/Model/Post_Estimation/Uniqueness`

File: `post_estimation_CLR_7.m`

- **Description.** This code computes the equilibrium for various random starting guesses for the wage function to confirm the equilibrium we converge to is unique.
- **Inputs.** Estimation outputs from baseline (`outputs_present_restud.mat`).
- **How to run.** Update the `replication_location` at the beginning of `post_estimation_CLR_7.m` using the working directory of the replicator, and run this file.
- **Outputs.** FIGURE: O.6.; stored in `Post_estimation/Uniqueness/Output`.

Folder `Estimation/Model/Heterogenous_Home_Production`

File: `hhp_CLR_0.m`

- **Description.** This code reproduces our model estimations that allow for heterogeneity in home production parameters across couple types (by presence of children and by skill composition of couples).
- **Inputs.** `inputs_present_types.mat`, `inputs_past_types.mat`, `inputs_present_ch0.mat`, `inputs_present_ch1.mat`, `inputs_past_ch0.mat`, `inputs_past_ch1.mat` (created in `inputs.m` in `Estimation/Data/data_inputs_for_estimation`; see Section D. of the Readme file for details.)
- **Auxiliary codes.** Uses the functions stored in folder `Heterogeneous_Home_Production/Functions` and its sub-folders.
- **How to run.** Update the `replication_location` at the beginning of file `hhp_CLR_0.m` using the working directory of the replicator, and run this file.
- **Outputs.** `outputs_present_ch0_restud.mat` and `outputs_present_ch1_restud.mat` (stored in `Heterogeneous_Home_Production\Functions\Present_child\Output`); `outputs_past_ch0_restud.mat` and `outputs_past_ch1_restud.mat` (stored in `Heterogeneous_Home_Production\Functions\Past_child\Output`); `SE_ch0_pr_restud.mat`, `SE_ch0_pa_restud.mat`, `SE_ch1_pr_restud.mat` and `SE_ch1_pa_restud.mat` (stored in `Heterogeneous_Home_Production\Functions\SE_child\Output`); `outputs_present_types_restud.mat` (stored in `Heterogeneous_Home_Production\Functions\HP_types\Output`); `outputs_past_types_restud.mat` (stored in `Heterogeneous_Home_Production\Functions\HP_types_past\Output`); `SE_hp_pr_restud` and `SE_hp_pa_restud.mat` (stored in `Heterogeneous_Home_Production\Functions\SE_types\Output`).

File: `hhp_CLR_1.m`

- **Description.** This code produces tables and figures of estimation results and model validation for the estimated model specifications that allow for heterogeneity in home production parameters across couple types (by presence of children and by skill composition of couples).
- **Inputs.** Estimation outputs from each specification in the baseline and past periods (produced in file `hhp_CLR_0.m`): `outputs_present_ch0_restud.mat`, `outputs_present_ch1_restud.mat`, `outputs_past_ch0_restud.mat`, `outputs_past_ch1_restud.mat`, `SE_ch0_pr_restud.mat`, `SE_ch0_pa_restud.mat`, `SE_ch1_pr_restud.mat`, `SE_ch1_pa_restud.mat`, `outputs_present_types_restud.mat`, `outputs_past_types_restud.mat`, `SE_hp_pr_restud`, `SE_hp_pa_restud.mat`.
- **How to run.** Update the `replication_location` at the beginning of the file `hhp_CLR_1.m` using the working directory of the replicator, and run this file.
- **Outputs** TABLE: O.17, O.18. FIGURES: O.14, O.15.; stored in `Heterogeneous_Home_Production\Output`.

D. Instructions for Producing Data Inputs for Structural Estimation and Post-Estimation

Main Folders

`Estimation/Data` and `Estimation/Dofiles`

Description

To produce all the intermediate data inputs for the structural estimation and post-estimation exercises, the following do-files (from folder `Estimation/Dofiles`) need to be run:

1. `CLR_estimation_1.do`
2. `CLR_estimation_2.do`

Note: The do-files `CLR_estimation_1.do` and `CLR_estimation_2.do` can only be run by accessing the SOEP Reanalysis Archive at DIW Berlin, where our working datasets are stored. Please see above for instructions on how to access that archive. However, there is no need to run these files if one is only interested in replicating the structural estimation analysis; see the description of folder `Estimation/Model` in Section C. of this Readme file.

To use the data inputs produced above and produce the inputs for the structural estimation (see Section C. above) the following files need to be run (from folder `Estimation/Data/data_inputs_for_estimation`):

1. `moments_and_weighting.m`
2. `inputs.m`

List of Programs and Produced Output

Folder `Estimation/Dofiles`

File: `CLR_estimation_1`

- **Description.** Produces data files that will be used to produce the inputs for the structural estimation and model validation (including targeted and untargeted data moments, bootstrapped moments, and confidence intervals of the data moments for several versions of the model—baseline present and past; heterogeneity in home production present and past; and heterogeneity by presence of children, present and past). The file is organized in three parts (each of them corresponding to a bookmark in the do-file):
  - PART I (first bookmark): Data Moments, produces all data moments.
  - PART II (second bookmark): Bootstrapped moments and confidence intervals.
  - PART III (third bookmark): Untargeted moments.
- **Inputs.** `SOEP_estimation_household.dta` from folder `Estimation/Data`
- **How to run:** To run the do-file, update the global `path_of_replicator` at the beginning of `estimation_CLR_1.do`, using the working directory of the replicator.
- **Outputs.** The outputs of running this do-file are:
  - PART I: `data_moments_baseline`, `data_moments_past`, `data_moments_baseline_types`, `data_moments_past_types`, `data_moments_baseline_child0`, `data_moments_baseline_child1`, `data_moments_past_child0`, `data_moments_past_child1`; stored in `Estimation/Data/moments`
  - PART II: `boot_baseline`, `ci_baseline`, `boot_past`, `ci_past`, `boot_baseline_child_0`, `ci_baseline_child_0`, `boot_baseline_child_1`, `ci_baseline_child_1`, `boot_past_child_0`, `ci_past_child_0`, `boot_past_child_1`, `ci_past_child_1`, `boot_baseline_types`, `ci_baseline_types`, `boot_past_types`, `ci_past_types`; stored in `Estimation/Data/moments`.
  - PART III: `data_moments_t3`, `data_moments_t4`, `data_moments_f7`; stored in `Estimation/Data/validation`.

File: `CLR_estimation_2`

- **Description.** Produces 4 auxiliary data files that will be used as inputs for the post-estimation and validation exercises.
- **Inputs.** `SOEP_estimation_individual.dta` (stored in `Estimation/Data`) and `SOEP_dataset` (stored in `Empirics/Data`).
- **How to run.** To run this do-file, update the global `path_of_replicator` at the beginning of `estimation_CLR_2.do`, using the working directory of the replicator.
- **Outputs.** Stores 4 auxiliary data inputs: `data_input_fig5_men`, `data_input_fig5_women`, `data_input_fig6_l`, `data_input_fig6_r`, used as inputs to produce Figures 5 and 6 from the paper; results are saved to `Estimation/Data/validation`.

Folder `Estimation/Data/data_inputs_for_estimation`

File: `moments_and_weighting.m`

- **Description.** This code creates the data inputs necessary for estimation of all model specifications: data moments and weighting matrices for estimation.
- **Inputs.** Data files from `Estimation/Data/data_inputs_for_estimation`
- **How to run.** Update the `replication_location` at the beginning of `moments_and_weighting.m` using the working directory of the replicator, and run this file.
- **Outputs.** Data moments and weighting matrices used for estimation; stored in `Estimation/Data/data_inputs_for_estimation/Output` and used in `Estimation/Model` (Section C. above).

File: `inputs.m`

- **Description.** This file saves the inputs for estimation of all model specifications.
- **Inputs.** Output files from `Estimation/Data/data_inputs_for_estimation/moments_and_weighting.m`. As a starting point for our estimations, we use the output from a global estimation routine saved in `outputs_ga_child0_past.mat`, `outputs_ga_child0`, `outputs_ga_child1_past`, `outputs_ga_child1`, `outputs_ga_types_past`, `outputs_ga_types`.
- **How to run.** Update the `replication_location` at the beginning of `inputs.m` using the working directory of the replicator, and run this file.
- **Outputs.** Creates all datasets used as inputs for estimation; stored in `Estimation/Data/data_inputs_for_estimation/Output` and used in `Estimation/Model` (Section C. above).

References

- German Time Use Survey (GTUS) Data for 1991/92 and 2012/13. German Federal Statistical Office. doi: 10.21242/63911.1992.00.00.4.1.0 and 10.21242/63911.2013.00.00.4.1.0
- A. Hall, A. Seifer, and M. Tiemann (2020). Bibb/baua employment survey of the working population on qualification and working conditions in germany 2012. Version 6.0. doi:10.7803/501.12.1.1.60, Research Data Center at BIBB (ed.); GESIS Cologne (data access); Bonn: Federal Institute for Vocational Education and Training.
- D. Rohrbach-Schmidt and A. Hall (2020). Bibb/baua-erwerbstaetigenbefragung 2012. BIBB-FDZ Daten-und Methodenberichte 1/2013 Version 6.0., Bonn: BIBB. ISSN 2190-300X.
- Socioeconomic Panel (2019) Data for Years 1984-2018, version 35, soep. doi:10.5684/soep.v35.
- A. Stuckemeier and C. Kuehnen (2013). German time use survey 2012/13. Technical report, Federal Statistical Office, Bonn, Germany.