

Replication Package for "Inequality and Income Dynamics in Germany"

Drechsel-Grau, Peichl, Schmid, Schmieder, Walz and Wolter

June 30, 2022

This replication package contains a supplemental appendix which is referred to the paper and all program files (Stata do-/ado-files, julia jl-files) that are used for the data pre-processing, preparation and to generate the main results. Those are provided in several zip-folders and shortly described in the following.

Contents

| | |
|---|-----------|
| Data Access | 2 |
| Requirements and Set-Up | 3 |
| 1 Data Pre-Processing and Data Preparation | 4 |
| 1.1 Prepare IAB data (IEB sample 2001–2016 and 1993–2018, SIAB 1985–2018 | 4 |
| 1.2 Prepare TPP data for the TPP sample 2001–2016 | 4 |
| 2 Combined IAB-TPP sample 2001–2016: Join Data and Produce Results | 5 |
| 2.1 IAB main sample input and descriptives: Earnings distributions groupwise summary and time series statistics | 5 |
| 2.2 Create weights from IAB summary statistics | 5 |
| 2.3 Reweighting of TPP Data | 6 |
| 2.4 Compute Statistics in TPP (Core Analysis on Earnings) | 6 |
| 2.5 Combine Earnings Results of IAB and TPP | 7 |
| 2.6 Compute Statistics in TPP (Analysis on Total Incomes) | 7 |
| 3 Generate Main Figures and Tables for the Paper | 8 |
| 3.1 Main Results for Joint and Separate Samples | 8 |
| 3.2 Reweighting Analyses | 8 |
| 4 Results for IAB Samples 1985–2018 and 1993–2018 | 10 |
| 5 Generate CSVs for GRID website | 11 |
| 5.1 Generate CSVs for GRID website: combined IAB-TPP data | 11 |
| 5.2 Generate CSVs for GRID website: mobility CSV topcoded (IAB only) | 11 |
| 5.3 Generate CSVs for GRID website: IAB samples 1985–2018 and 1993–2018 | 11 |

Data Access

Access to Micro Data from IAB:

Both administrative panel datasets - the IEB V13.01.01-190111¹ and the SIAB 7519_v1² - on individuals' characteristics and wages used for (aggregate) data linkage and the analyses are collected, processed and kept by the IAB (IEB V13.01.01-190111) and the Research data Center (FDZ) of the Federal Employment Agency (BA) at the IAB (SIAB 7519_v1). The data is legally classified as "social data" and stored by the IAB according to the German Social Code, Book III (Sozialgesetzbuch, III. Buch). The data contains sensitive information and is therefore subject to the confidentiality regulations of the German Social Code (Book I, Section 35, Paragraph 1). The data is held by the Institute for Employment Research (IAB), Regensburger Str. 104, D-90478 Nürnberg, email: iab@iab.de, phone: +49 911 1790. If you wish to access this data for replication purposes please also see the statement on data access for replication purposes by the IAB on its website <https://www.iab.de/en/daten/replikationen.aspx> and on the FDZ website https://fdz.iab.de/en/FDZ_Data_Access.aspx.

Access to Micro Data of the TPP from the Federal Statistical Office:

The German Taxpayer Panel (TPP) is available to researchers via The Research Data Centres of Official Statistics in Germany which consist of the Research Data Centre of the Federal Statistical Office and the Research Data Centre of the Statistical Offices of the Federal States. See <https://www.forschungsdatenzentrum.de/de/steuern/tpp> for more information (albeit only available in German) on the TPP data.

In general, the use of official microdata is regulated by law (Gesetz über Steuerstatistiken (StStatG) in case of the TPP), committed to confidentiality, bound to a specific user or group of users, bound to a specific purpose, temporary, contractually agreed, and liable for costs. Details can be found here: <https://www.forschungsdatenzentrum.de/en/terms-use>. The standard user charge is 250 Euro per statistic, survey year and way of access. To gain access to TPP data, a researcher has to fill out an application form (<https://www.forschungsdatenzentrum.de/en/request>). The TPP is then available via controlled remote data processing (kontrollierte Datenfernverarbeitung, KDFV). If you wish to access this data for replication purposes please get in contact with the authors.

¹IEB: Integrated Employment Biographies. This is process data from the Federal Employment Agency held at the IAB.

²IEB: Sample of Integrated Employment Biographies. This is a scientific data product containing a 2% sample of the IEB in a pre-processed format.

Requirements

This requirements and set-up instructions hold for all steps below if not stated differently.

- **Software.** Stata Version 16.0 or later. The code was created using Stata 16.1 and 17.0. Some code uses commands by state of Stata 13. The julia code was written for julia 1.6.2. A `Project.toml` and `Manifest.toml` file is in the zip-folder.
- **Self-written and frequently used ado-files.** The following ado-files are frequently used and need to either be downloaded manually and to be placed in the `prog` folder or to be installed using `ssc install`:
 - `gtools`
 - `labmask`
 - `outreg2`
 - `stetig`
 - `unique`
 - `winsor`
 - `inequal7`

We use a self-written ado-file that is provided in the zip-folder `ado_file.zip` (incl. a help-file) and those need to be placed in the `prog` folder:

- `cellgraph.ado`
- `cellgraph.hlp`

1 Data Pre-Processing and Data Preparation

1.1 Prepare IAB data (IEB sample 2001–2016 and 1993–2018, SIAB 1985–2018)

- Stata do-files (from the zip-folder `do-files-data-prep-IAB` in this replication package):
 - `master_preparation_IEB.do`
 - `master_preparation_SIAB.do`
 - `09_wages_imputation.do`
- Input data:
 - `ieb_10perc_clean.dta`: 10% pre-cleaned random sample of raw IEB data, version January 11, 2019
 - `SIAB_7519_v1.dta`: SIAB version 1975–2019, Apr 16 2021
 - `SIAB_7519_v1_bhp_basis_v2.dta`: base establishment file for SIAB version 1975–2019, Apr 16 2021
- Output data:
 - `clean_yearly_data.dta`: annual individual-level earnings dataset of 10% random sample of the IEB
 - `clean_yearly_siab.dta`: annual individual-level earnings dataset of SIAB (2% pre-cleaned IEB sample)

1.2 Prepare TPP data for the TPP sample 2001–2016

- Stata do-files (from the zip-folder `do-files-TPP-kdfv-analysis` in this replication package):
 - `000_master_fdz.do`
 - `000_master_fdz_preamble.do`
 - `001_Initialize.do`
 - `101_DataPrep.do`: This do-file draws a 25 sample of all units in the raw TPP data and converts the tax-unit-level data to an individual-year panel.
 - `102_ImputeCivilServants.do`: This do-file imputes the indicator variable for civil servants (or social security workers) before 2010 to correct a break in the variable from 2009 to 2010.
- Input data:
 - The research data center of the Federal Statistical Office select the following 25 variables from the raw TPP data (full population, 2001-2016) and convert the data to a Stata data file (`yourrawtppdata.dta`): `ida`, `idb`, `ef1`, `ef8`, `ef11`, `ef12`, `ef18`, `ef43`, `ef44`, `c65163`, `c65164`, `c65121`, `c65122`, `c65221`, `c65222`, `c65101`, `c65102`, `c65241`, `c65242`, `c65141`, `c65142`, `c65261`, `c65262`, `c47016`, `c47116`, `c48016`, `c48116`.
- Output data:
 - `BASEDATA_vollmaterial25.dta`

2 Combined IAB-TPP sample 2001–2016: Join Data and Produce Results

2.1 IAB main sample input and descriptives: Earnings distributions groupwise summary and time series statistics

- Stata do-files (from the zip-folder `do-files-IAB-main-sample` in this replication package):
 - `master.IAB_main_sample_stats.do`
 - `0_Initialize.do`
 - `1_Gen_Base_Sample.do`
 - `1_Gen_Base_Sample_Dist.do`
 - `2_DescriptiveStats.do`
 - `3_Inequality.do`
 - `4_Volatility.do`
 - `5_Mobility.do`
 - `7_Paper_Figs_long.do`
 - `7_Paper_Figs_short.do`
 - `200_data_privacy_restrictions.do`
 - `200_myprogs.do`
 - `201_Initialize.do`
 - `206_Distributions_iab_hw.do`
 - `myplotsettings.do`
 - `myplots.do`
 - `myprogs.do`
- Input data:
 - `clean_yearly_data.dta`
- Output data (in 'out' folder):
 - `sumstats_labor_maleage.csv` in folder '`date Descriptive_Stat - short`': Earnings statistics for main sample IAB.
 - Several CSV files in folders '`inequality`', '`volatility`' and '`mobility`') in directory `date 206_Distributions`: Percentiles and other distributional measures of earnings variables on inequality, volatility and mobility of main sample in IAB data.
 - `data_diff_dep_var_cohort_plot_v2.csv` in folder '`date Data ddv cohort plot`': Cohort-wise relative trends in demographic characteristics.
 - `transition_probabilities_jobtype1.csv` in folder '`date Transition probs jtl`': Trends in probabilities of transitions between full-time, part-time and mini-job.

2.2 Create weights from IAB summary statistics

- Julia jl-files (from the zip-folder `julia-files-combine-IAB-TPP` in this replication package):
 - `000_master.jl`
 - `100_iab_earnings_weights.jl`: convert year-gender-age-specific earnings percentiles to a discrete distribution (bins) to be used to compute the weights for the TPP data.

- Input data:
 - PATH TO IAB RESULTS/*date* 206_Distributions/inequality/sumstats.labor.maleage.csv
- Output data:
 - `iab_weights_new.csv`: Number of observations in different earnings bins by year, gender, age. This file has to be transferred to the FDZ-TPP (separate application needed for linking external data to the TPP) and stored in the folder specified under the global variable `dta_folder` in the master-do-file `000_master_fdz.do`.

2.3 Reweighting of TPP Data

- Stata do-files (from the zip-folder `do-files-TPP-kdfv-analysis` in this replication package):
 - `000_master_fdz.do`
 - `000_master_fdz_preamble.do`
 - `001_Initialize.do`
 - `103_CorrectSampleSelection.do`
- Input data:
 - `BASEDATA_vollmaterial25.dta`
 - `iab_weights_new.csv`
- Output data:
 - `BASEDATA_vollmaterial25.dta`

2.4 Compute Statistics in TPP (Core Analysis on Earnings)

- Stata do-files (from the zip-folder `do-files-TPP-kdfv-analysis` in this replication package):
 - `000_master_fdz.do`
 - `000_master_fdz_preamble.do`
 - `001_Initialize.do`
 - `106_Descriptives.do`
 - `201_Gen_Base_Sample.do`
 - `206_Distributions.do`
- Input data:
 - `BASEDATA_vollmaterial25.dta`
- Output data:
 - several CSV files in folders '`inequality`', '`volatility`' and '`mobility`' in directory `date 206_Distributions` (labor, rv)

2.5 Combine Earnings Results of IAB and TPP

- Julia jl-files (from the zip-folder `julia-files-combine-IAB-TPP` in this replication package):
 - `000_master.jl`
 - `101_tppiab_inequality.jl`
 - `102_tppiab_volatility_new.jl`
 - `103_tppiab_mobility_new.jl`
 - `functions_combine_iab_tpp.jl`
 - `functions_inequality.jl`
 - `functions_volatility.jl`
 - `functions_other.jl`
- Input data:
 - `PATH TO IAB-RESULTS/date 206_Distributions`
 - `PATH TO TPP-RESULTS/date 206_Distributions (labor, rv)`
- Output data:
 - Main results for core analysis on earnings inequality and dynamics: store them in sub-folders `'inequality'`, `'volatility'` and `'mobility'` in the directory `PATH TO JOINT-RESULTS`

2.6 Compute Statistics in TPP (Analysis on Total Incomes)

- Stata do-files (from the zip-folder `do-files-TPP-kdfv-analysis` in this replication package):
 - `000_master_fdz.do`
 - `000_master_fdz_preamble.do`
 - `001_Initialize.do`
 - `401_Gen_Base_Sample.do`
 - `406_Distributions.do`
- Input data:
 - `BASEDATA.vollmaterial25.dta`
- Output data:
 - Several CSV files in folders `'inequality'`, `'volatility'` and `'mobility'` in directory `date 406_Distributions (totinc.noKV)`

3 Generate Main Figures and Tables for the Paper

3.1 Main Results for Joint and Separate Samples

- Stata do-files (from the zip-folder `do-files-postanalysis` in this replication package):
 - `master_post.do`
 - `001_initialize.do`
 - `200_descriptives.do`
 - `200_imputation_civilservants_tpp.do`
 - `201_core_inequality.do`
 - `202_core_volatility.do`
 - `203_core_mobility.do`
 - `204_core_reweighting.do`
 - `301_income_sources.do`
 - `301_total_inequality.do`
 - `301_total_volatility.do`
 - `301_total_mobility.do`
 - `myplotsettings.do`
 - `programs_figures.do`
 - `programs_general.do`
 - `programs_tables.do`
- Input data:
 - csv-files in `PATH TO JOINT-RESULTS`
 - csv-files in `PATH TO TPP-RESULTS`
 - csv-files in `PATH TO IAB-RESULTS`
- Output data:
 - tex-files (tables) and pdf-files (figures) in sub-folders `'data'`, `'core'`, `'specific'` and `'app-tppiab'` in the folder `PATH TO FIGURES-TABLES`

3.2 Reweighting Analyses

- Stata do-files (from the zip-folder `do-files-reweighting-analyses` in this replication package):
 - `master_IAB_reweigthing_analyses.do`
 - `0_initialize.do`
 - `1_Gen_Base_Sample_rw.v2.do`
 - `3_Inequality_rw.v2.do`
 - `7_Paper_Figs_short_rw.do`
 - `myplots.do`
 - `myplotsettings.do`
 - `myprogs.do`
- Input data:
 - `clean_yearly_data.dta`: Annual individual-level earnings data from the IAB (10%-IEB sample), using imputed wages.

- Output data:
 - `reweights_rvlX.clean_yearly_data.dta`, with $X \in [1, 2, 3, 5b, 6, 7]$: Weights for re-weighting w.r.t. variable X based on the `clean_yearly_data` dataset.
 - `L_logearn_male_sumstat.csv` and `L_logearn_sumstat.csv`: Summary stats (by gender) for each re-weighting analysis in the respective folder `date Inequality - short_rwX` (again with $X \in [1, 2, 3, 5b, 6, 7]$, indicating different variables to re-weight to).

4 Results for IAB Samples 1985–2018 and 1993–2018

- Stata do-files (from the zip-folder `do-files-IAB-long-samples` in this replication package):
 - `master.IAB.long.samples.do`
 - `0_Initialize.do`
 - `1_Gen.Base.Sample.do`
 - `3_Inequality_P85.do`
 - `4_Volatility_P85.do`
 - `7_Paper_Figs_P85.do`
 - `myplots.do`
 - `myplotsettings.do`
 - `myprogs.p85.do`
- Input data:
 - `clean_yearly_data.dta`
 - `clean_yearly_siab.dta`
- Output data:
 - Several summary and volatility statistics files in CSV format in respective folders `date inequality`, `date volatility` and `date mobility` used for plotting several figures.

5 Generate CSVs for GRID website

5.1 Generate CSVs for GRID website: combined IAB-TPP data

- Stata do-files (from the zip-folder `do-files-grid` in this replication package):
 - `Stats_GER.do`
 - `Density_GER.do`
 - `Rank_GER.do`
 - `Mobility_GER.do`
- Input data:
 - csv-files in `PATH TO JOINT-RESULTS`
 - csv-files in `PATH TO TPP-RESULTS` (for Pareto tails)
- Output data:
 - `Stats_GER.csv`
 - `Density_GER.csv`
 - `Rank_GER.csv`
 - `Mobility_GER.csv`

5.2 Generate CSVs for GRID website: mobility CSV topcoded (IAB only)

- Stata do-files (from the zip-folder `do-files-IAB-only-CSVs` in this replication package):
 - `master_IAB_samples_CSVs.do`
 - `_1_prep_datatsets.do`
 - `_2_data_prep.do`
 - `_4_mobility_csv.do`
 - `0_Initialize_nonimp.do`
 - `1_Gen_Base_Sample_nonimp.do`
 - `myprogs.do`
- Input data:
 - `clean_yearly_data.dta`
 - `ieb_10perc_clean.dta`: 10% pre-cleaned random sample of raw IEB data, version January 11, 2019
- Output data:
 - `mobility_GER_IABsample_topcoded.csv`: Mobility CSV for IAB sample 2001–2016, topcoded (majority rule) as separate category.

5.3 Generate CSVs for GRID website: IAB samples 1985–2018 and 1993–2018

- Stata do-files (from the zip-folder `do-files-IAB-only-CSVs` in this replication package):
 - `master_IAB_samples_CSVs.do` [generates all long sample CSVs]
 - 3 folders for the 3 samples:
 - `csv_sample19852018`

csv_sample19932018

csv_sample20012016

- Each folder contains the (non-exchangeable) similarly named do-files:

0_Initialize.do

1_Gen_Base_Sample.do

2_DescriptiveStats.do

3_Inequality.do

4_Volatility.do

5_Mobility.do

6_Insheeting_datasets_reduced.do

data_protect.do

myplots.do

myprogs.do

- Input data:
 - clean_yearly_data.dta: Needed for IAB samples 2001–2016 and 1993–2018.
 - clean_yearly_siab.dta: Needed for IAB sample 1985–2018.
- Output data:
 - Stats_GER_IAB1985-2018.csv
 - Density_GER_IAB1985-2018.csv
 - Stats_GER_IAB1993-2018.csv
 - Density_GER_IAB1993-2018.csv
 - Stats_GER_IAB2001-2016.csv
 - Density_GER_IAB2001-2016.csv