Data and Code for: The Long Run Evolution of Absolute Intergenerational Mobility

Yonatan Berman

All MATLAB and R codes, as well as data files, are available in **openicpsr-141761.**

The replication repository contains:

1. All code files needed to produce the analysis
2. A script that reproduces all figures
3. Data (as detailed below)

The paper is based mainly on income data from the World Inequality Database (WID). Additional data, on rank correlations in different countries, was taken from published papers as described in the paper (in Appendix F).

To fully replicate the paper results using WID data, it is necessary to use an online interface (https://wid.world/gpinter/). This README, in addition to describing the content of the repository, will therefore include two guides. The first guide is on how to use the online interface to produce datafiles that should be used to replicate the results. The second guide is a demonstration of the codes used in the paper and included in the repository that produce absolute mobility estimates.

The paper relies crucially on the GPINTER interface and on the WID data. It is thus important to clarify that the author is not the maintainer/author of the software nor of the World Inequality Database and has no control over the availability or implementation of the software package and the database.

Software Requirements:

- MATLAB (version R2020a and above)
- R (the code was run on R-3.5; devtools package required)
- GPINTER (online interface, see above)
- The code was last run on an Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz, 1801 Mhz with 8.00 GB of installed memory (RAM) with Windows 10 Home.
- Time requirements: all figure scripts typically take about 5 minutes to run in total. Some of the underlying calculations could take up to several days to run (see specific documentation in the codes).

Repository content:

The repository includes three folders:

- Figures
- Other code
- Other data

Figures:

The Figures folder contains the MATLAB scripts and workspaces (a MATLAB compatible data format) which reproduce all figures in the main text and the online appendix. `MS_master_figures.m` is a script that runs all figure scripts, each produces one figure. The figures are exported locally as jpg files. The values in Tables H.1 and H.2 are produced with Figures C.4 and H.1, respectively.

No workspace is needed for running `figure1.m` and `figure3.m`. The other figure scripts use the provided workspaces as follows:

- Figure 2: `copulas_measures.mat` and `figure2_results.mat`
- Figure 4: `main_results.mat`

- Figure 5: `counterfactual_results.mat`
- Figure 6: `counterfactual_results_US_early.mat`
- Figure 7: `FR_inequality_measures_results.mat`
- Figure C.1: `decomposition_check.mat`
- Figure C.2: `copula_moves.mat`
- Figure C.3: `models.mat`
- Figure C.4 (+ Table H.1): `mobility_with_err.mat`
- Figure C.5: `overturn.mat`
- Figure C.6: `limits.mat`
- Figure D.1: `copulas.mat`
- Figure E.1: `assortative_check.mat`
- Figure E.2: `france_equal_indiv.mat`
- Figure E.3: `fr_us_ages.mat`
- Figure E.4: `cps_res.mat`
- Figure E.5: `baseline_labor.mat`
- Figure E.6: `pre_post_res.mat`
- Figure E.7: `tweaks_res.mat`
- Figure E.8: `tweaks_over_time.mat`
- Figure G.1: `mobility_comparison.mat`
- Figure H.1 (+ Table H.2): `counterfactuals.mat`

This folder includes additional functions and scripts called by the figure scripts above:

- `gaussian_rhoc.m` – converts Gaussian copula parameter to rank correlation.
- `absmob_logn.m` – the basic function that implements the bivariate log-normal computes the measure of absolute mobility given two vectors (of parents and children).
- `mobility_decompose.m` – a simple function that uses the counterfactual calculation results to decompose the total change in absolute mobility to the contribution of income growth and change in the inequality. It is called by `figureH1.m` and used to produce Table H.2.

- `moving_average.m` – a simple moving average smoother used in Figure E.4.

Other code:

- `absmob_counterfactuals_script.m` – a script that imports income data series, calculates absolute mobility for the given series, as well as counterfactual scenarios (see Section III.C) and plots the results (see also guide below).
- `absmob_series_script.m` – a script that imports income data series, calculates absolute mobility for the given series, and plots the results (see also guide below).
- `absmob_series_script_with_copula.m` – a script that imports income data series and a copula (represented by a matrix), calculates absolute mobility for the given series data with the given copula, and plots the results (see also guide below).
- `absmob_with_copula.m` – computes the measure of absolute mobility given two vectors (of parents and children) and given a discrete copula (a matrix).
- `absmob0.m` – the basic function that computes the measure of absolute mobility given two vectors (of parents and children).
- `check_copula_vs_analytic.m` – this script creates the results presented in Figure C.1 in the appendix. It assumes a log-normal bivariate distribution for the joint income distribution of parents and children based on parameters for France (taken from `France_Data_For_Decomposition.mat`, is also in the folder), and then simulates discrete copulas at different resolutions to check whether the analytic result for the absolute mobility is recovered when discrete copulas are used.
- `clayton_rho.m` – converts Clayton copula parameter to rank correlation.
- `clayton_theta.m` – converts rank correlation to Clayton copula parameter.
- `couple_vecs.m` – gets two vectors of the same size, as well as a parameter and a type of copula (as a string) and returns the vectors with their joint rank distribution represented by the copula model specified.
- `couple_vecs_plackett.m` – gets two vectors of the same size, as well as a Plackett copula parameter and returns the vectors with their joint rank distribution represented by the Plackett copula parameterized with the given parameter.

- `gaussian_rhoc.m` – converts Gaussian copula parameter to rank correlation (also in Figures folder).
- `gaussian_theta.m` – converts rank correlation to Gaussian copula parameter.
- `gumbel_rho.m` – converts Gumbel copula parameter to rank correlation.
- `gumbel_theta.m` – converts rank correlation to Gumbel copula parameter.
- `import_WID_for_GPINTER.R` – installs package, imports, downloads, and exports distributional data from the World Inequality Database.
- `plackett_rho.m` – converts Plackett copula parameter to rank correlation.
- `plackett_rnd.m` – generates bivariate random numbers from the Plackett copula.
- `plackett_theta.m` – converts rank correlation to Plackett copula parameter.
- `run_cps_age_abs_weights.m` – this script creates the main results presented in Figure E.4 in the appendix. It estimates absolute mobility for the United States given multiple adjustments to income definition and unit of observation using the CPS data (taken from `CPS_workspace.mat`). This script makes use of three additional scripts in the folder: `standardize_weight.m`, `make_hincome1.m` and `make_hincome2.m`, and a MATLAB workspace: `theta_rankcorr.mat`.
- `write_excel_for_GPINTER.m` – reads data downloaded from WID using the R script `import_WID_for_GPINTER.R` and converts this file into an xlsx file in a format that can be read by the GPINTER online interface.

Other data:

- `copula_example.csv` – a 10x10 copula that represents the joint intergenerational income rank distribution in the United States as estimated in the paper: The fading American dream: Trends in absolute income mobility since 1940, by R. Chetty et al. (https://doi.org/10.1126/science.aal4617) for the 1980 birth cohort.
- `RankCorrelations.csv` – a table which includes the rank correlation values used for the absolute mobility calculations (see also Appendix F in the paper).
- `samples.csv` – a table that contains simulated samples that are based on parameters which represent the WID data for the years 1915 – 2014 in France. It is an example used by the MATLAB scripts described in Guide 2 below.

- `theta_rankcorr.mat` – see 'Other code'.
- `theta_rankcorr_clayton.mat` – a MATLAB workspace to allow converting between rank correlations and the Clayton copula parameter.
- `theta_rankcorr_plackett.mat` – a MATLAB workspace to allow converting between rank correlations and the Plackett copula parameter.
- `WID_data.csv` – an output example of data downloaded from WID and exported using R (`import_WID_for_GPINTER.R`), for the purpose of using the GPINTER online interface. See Guide 1 below for more details.
- `WID_input_for_GPINTER.xlsx` – an input example for the GPINTER online interface produced using the MATLAB script `write_excel_for_GPINTER.m`, based on `WID_data.csv`. See Guide 1 below for more details.

Guide 1: Producing simulated samples using GPINTER online interface:

To replicate the results in the paper, it is necessary to obtain simulated samples of the income distribution in different countries in different years. These samples rely on data from the World Inequality Database (WID). Creating these samples was done using an online interface called GPINTER (https://wid.world/gpinter/), named after the generalized Pareto interpolation method. The method itself is described in detail in the paper: Generalized Pareto Curves: Theory and Applications, by T. Blanchet, J. Fournier, and T. Piketty (https://doi.org/10.1111/roiw.12510).

This guide will describe the necessary steps to produce simulated samples in GPINTER using data from WID. The relevant code mentioned in this guide is included in the 'Other code' folder.

Step 1: Downloading data from WID using R

WID has both Stata and R supported packages. For full documentation on the R package please refer to https://thomasblanchet.fr/wp-content/uploads/2020/04/wid-vignette.pdf.

To create a simulated sample using the GPINTER interface we need to describe for each case (case can be a single country in some year, say France in 1980) the following:

- Year

- Average income
- The threshold to be included in rank *p*
- The top average income above rank *p*

This information, for at least four ranks *p* is required to use the interpolation method which is used to create the samples, as well as the average income across the population. Whenever possible we use information on 14 ranks: (*p*=0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99, 0.999, 0.9999, 0.99999).

The R script `import_WID_for_GPINTER.R` in the 'Other code' folder is a short script that downloads these data (the current example is for France in the years 1950 – 2000, but it can be easily tweaked), and exports it to the file `WID_data.csv` (example is in the 'Other data' folder).

There are additional parameters included in the R commands, such as age, income concept, and population type, which can be changed (subject to their availability in the database). Refer to the full documentation of the WID R tool in https://github.com/WIDworld/wid-r-tool.

Step 2: Preparing the data for GPINTER

The next step is to rewrite the WID data to make it compatible with the format required by the GPINTER tool. This is done using the MATLAB script `write_excel_for_GPINTER.m` (see 'Other code' folder). This script converts the output of the R script (`WID_data.csv`) into an Excel file (`WID_input_for_GPINTER.xlsx`) that will be uploaded to the online interface.

Step 3: Running GPINTER

The Excel file created using the MATLAB script can be now read by GPINTER. In https://wid.world/gpinter/, under 'Import file(s)' browse and choose the Excel file created in the previous step, then make sure all sheets are selected in the pop-up window and click 'Import selected'. Hit 'Run'. When GPINTER is done hit 'Close'. At this point the GPINTER output can be plotted and exported in various ways.

Click 'SAMPLE', set the sample size (default is 1000000) and hit 'Download as CSV'. The

downloaded file is a table, and each column consists of a simulated sample for one year, based on the downloaded data from WID. These samples can now be used for producing absolute mobility estimates and the rest of the calculations in the paper.

Guide 2: Producing absolute mobility estimates and additional calculations:

This complements the previous guide. The first step for fully replicating the paper results is obtaining the simulated samples as described in Guide 1. The next step is using the samples for estimating absolute mobility, as well as to produce additional relevant calculations, such as the counterfactual scenarios described in the paper in Section III of the paper.

We describe below three types of calculations. In all of them we assume that a file `samples.csv` is a table and each column consists of a simulated sample for one year, based on the downloaded data from WID. The first row in the table consists of the year numbers. The file `samples.csv` in the 'Other data' folder is an example of such file in the right format. All the MATLAB codes described below are included in the folder 'Other code', and the necessary MATLAB workspaces required for the full functionality of the codes are in 'Other data'.

Calculation 1: Estimating a series of absolute mobility estimates with a given rank correlation

The first calculation is the basic calculation which the paper is built on – given income samples that span over a time period starting at year X, we match together the samples in years X and X+30, X+1 and X+31, X+2 and X+32, and so on, using a modeled copula with the given rank correlation. We then calculate the absolute mobility that results in the series of matched sample pairs.

The MATLAB script `absmob_series_script.m` implements this using an example data file `samples.csv` and plots the results.

By default, `absmob_series_script.m` uses a Gumbel copula to match the samples together. This can be replaced by other copula models supported by the `copularnd` command in MATLAB. The relevant ones are Gaussian and Clayton copulas. The relevant data files to use other copula models are included in 'Other data'.

Calculation 2: Producing counterfactual scenarios

The next calculation produces, in addition to the baseline absolute mobility series, two counterfactual calculations. In the first counterfactual calculation, the distribution shape remains the same as in the first sample in a given dataset. It effectively calculates mobility as if inequality did not change. The second counterfactual scenario allows the distribution shape to change but assumes that the overall income growth, between the first and last samples, was spread evenly over time, i.e., the same growth rate in every year. This produces a calculation in which changes in income growth over time are neutralized. This is discussed in detail in Section III.C.

The MATLAB script `absmob_counterfactuals_script.m` implements this using an example data file `samples.csv` and plots the results. Same as before, this script uses a Gumbel copula to match the samples together.

Calculation 3: Estimating a series of absolute mobility estimates with a given copula

The third calculation is the same as Calculation 1 with a major difference – here the copula itself is an input of the calculation. It is not a modeled copula. This gives an additional important degree of freedom in the calculations.

The MATLAB script `absmob_series_script_with_copula.m` implements this using the example data file `samples.csv` and plots the results. It also makes use of an external copula, imported from `copula_example.csv`. The copula example is based on accompanying data of the paper: The fading American dream: Trends in absolute income mobility since 1940, by R. Chetty et al. (https://doi.org/10.1126/science.aal4617).

References:

1. Blanchet, Thomas, Juliette Fournier, and Thomas Piketty. "Generalized Pareto interpolation." Online interface. URL: https://wid.world/gpinter/. See also: Blanchet, Thomas. "Applying generalized Pareto interpolation with gpinter." URL: https://thomasblanchet.fr/wp-content/uploads/2020/04/gpinter-vignette.pdf.

2. Blanchet, Thomas, Juliette Fournier, and Thomas Piketty. "Generalized Pareto curves:

theory and applications." *Review of Income and Wealth* (2021). DOI: 10.1111/roiw.12510.

3. Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. "The fading American dream: Trends in absolute income mobility since 1940." Science 356, no. 6336 (2017): 398-406. DOI: 10.1126/science.aal4617.

4. The World Inequality Database. “Australia, Canada, Denmark, Finland, France, Japan, Norway, Sweden, UK and USA per Adult National Income.” URL: http://wid.world/data/, Access Date: Oct-10-2017.