

ABSOLUTE MOBILITY: REPLICATION READ ME

This document describes the steps to replicate our results. It is split into 2 parts. Part 1 describes how to assemble various datasets our code requires to run; Part 2 describes steps to run the code and generate the figures/tables in our paper.

PART 1: Assembling the Data

The data for our project come from multiple sources:

1. CPS and Census data, downloaded from IPUMS-CPS and IPUMS-USA, respectively
2. Copula relating parents & children by income-rank, generated at the IRS
3. Federal tax rates & exemptions from 1940-1959 from Wilson (2002)
4. Price deflators and real GDP from the Federal Reserve Economic Data database (FRED)
5. Estimates of near-cash and in-kind transfers from Fox et al. (2015)

We provide instructions on downloading #1 below. Files related to #2, #3, and #4 are provided in our zip folder. We thank Fox et al. (2015) for making #5 available to us; currently, we are unable to share that data ourselves.

1.1. Census Data from IPUMS-USA:

Download Census data from IPUMS-USA [\[link\]](#) for every decade from 1940 to 2010. You will need to create a (free) login ID to do so. For 1940 and 1950, you will need to download two versions of the data, one for the baseline analysis and one to compute the number of working-age families in the counterfactuals.

Table 1 lists the samples to select for each decade. Each sample must include the following variables: YEAR, SERIAL, STATEFIP, GQ, GQTYPE, PERNUM, FAMUNIT, FAMSIZE, NCHILD, MOMLOC, POPLOC, SPLOC, SEX, AGE, BPL, INCWAGE, FTOTINC, INCTOT. Each sample may require additional variables or selection criteria, also described in Table 1 below:

Table 1: Samples and variables to extract from IPUMS-USA

Decade	Sample	Additional Variables	Selection criteria
1940	1% Sample	SELFWTSL, RACE, OCC1950, CLASSWKR, INCNONWG	Select cases in the self-weighting sample. That is, before submitting your extract, check the box for SELFWTSL = 2 on the "Select Cases" page (including all persons residing in a household with a selected person).
1940	1% Sample		Counterfactuals sample: No criteria --- include everyone

1950	1% Sample	SELFWTSL, RACE, OCC1950, CLASSWKR	Select cases in the self-weighting sample. That is, before submitting your extract, check the box for SELFWTSL = 2 on the "Select Cases" page (including all persons residing in a household with a selected person).
1950	1% Sample		Counterfactuals sample: No criteria --- include everyone
1960	1% Sample		No criteria -- include everyone
1970	1% Form 1 Metro Sample	INCWELFR	No criteria -- include everyone
1980	1% Metro Sample	INCWELFR	No criteria -- include everyone
1990	1% Unweighted Sample	INCWELFR	No criteria -- include everyone
2000	1% Unweighted Sample	INCWELFR, INCSUPP	No criteria -- include everyone
2010	1% ACS Sample	INCWELFR INCSUPP HHWT	No criteria -- include everyone

Note that for the state-level analyses, we use the 1940 Full Population data and the 1980 5% sample for greater precision. Though the 1% samples listed above produce very similar results, you should download those larger samples if you desire full reproducibility. When downloading a sample, download both the data file and the associated .do file (which will have a name like USA_<downloadID>.do). The .do file applies labels to values and performs some cleaning of its own. Run the .do file, and save the resulting .dta file under the title CensusYEAR_raw.dta (for e.g., Census1940_raw.dta, Census1950_raw.dta etc). For 1940 and 1950, save the second version as CensusYEAR_raw_cf.dta. We will refer to the folder containing these files as "indir_raw", though you may name it anything you like.

1.2. CPS Data from IPUMS-CPS

Next, download CPS data from IPUMS-CPS [\[link\]](#) that includes all ASEC samples from 1968 to 2015. Unlike the cCensus case, download all years' data as a single file; as before, run the associated .do file to apply labels, and save the result as CPS_1968to2015.dta. The list of variables required (across years) is:

YEAR, SERIAL, PERNUM, AGE, SEX, BPL, MARST, NCHILD, FAMUNIT, FAMSIZE, MOMLOC, POPLOC, SPLOC, WTSUPP, HWTSUPP, GQ, FTOTVAL, INCALIM, INCALOTH, INCASIST, INCBUS, INCCHILD, INCDISAB, INCDIVID, INCDRT, INCEDUC, INCFARM, INCIDR, INCINT, INCGOV, INCLONGJ, INCOTHER, INCRENT, INCRETIR, INCSS, INCSSI, INCSURV, INCTOT, INCUNEMP, INCVET, INCWAGE, INCWELFR, INCWKCOM, OINCBUS, OINCFARM, OINCWAGE.

Next, download the .dta-file containing the income component rank proximity swap values for

1976-2010 [[link](#)]. Name it CPS_swapvalues.dta. We need this file to undo top-coding of income.

Place CPS_1968to2015.dta and CPS_swapvalues.dta in **indir_raw**

1.3. Zip Data

Place all the unzipped data files (the copula files, price deflators, etc.) in a separate folder. We'll refer to it as **indir_clean**, though you may name it whatever you like.

We're We are now ready to run the code.

Part 2: Running the code

Here i's an overview of our main scripts:

clean_up.do: cleans raw IPUMS data and reshapes it into a structure that the program absmob.ado can use

absmob.ado: Workhorse program -- computes marginal incomes and mobility statistics based on various input parameters

robustness.do: calls absmob.ado with various input parameters / specifications to generate results for robustness tests.

robustness_tax.do: Computes absolute mobility after taking federal taxes into consideration.

counterfactuals.do: Computes absolute mobility under counterfactual simulations.

draw_figs.do & **draw_app_figs.do**: uses datasets produced by robustness.do and counterfactuals.do to produce the paper's main figures (and appendix figures respectively).

The sequence to generate our results is as follows:

Prelim: Install the sStata module fastxtile [[link](#)]

Step 1: Edit & run **clean_up.do**.

Edit the global macro rawdataroot to point to the folder that contains the downloaded IPUMS data (**indir_raw**). Edit the global intermedroot to refer to **indir_clean**.

Run clean_up.do. It will clean and reshape the IPUMS data.

At the end of this step, **indir_clean** will contain the results of this run

(census<decade>_families.dta and census<decade>_couples.dta for each decade, and a single CPS_clean.dta), as well as the unzipped data files (copula, price deflators, etc.) that were already placed in it.

Step 2: Edit & run **robustness.do**

Edit the global macro absmob_indir to point to **indir_clean**. Set the macro absmob_sourcedir to the path of the folder that contains all the downloaded scripts. Set the macro absmob_resultsdir to point to a local directory that will contain the marginal incomes and mobility stats for all our baseline, robustness, and counterfactual results. Make sure it has adequate space (About 20 GB).

Run **robustness.do**. This will generate our baseline results and our robustness specifications by calling the program **absmob.ado** repeatedly with different parameters. Each specification will create a new subfolder within `absmob_resultsdir` and store the results in it.

Step 3: Edit & Run **robustness_tax.do**

`Robustness.do` produces all robustness specifications *except* those associated with our tax series. **robustness_tax.do** generates those results.

Edit the global macros `absmob_indir`, `absmob_sourcedir`, `absmob_resultsdir` as described in step #2.

robustness_tax.do uses NBER's `Taxsim` [[link](#)] to estimate taxes for years after 1959. To speed up computation, `Taxsim`'s authors have provided us with a copy of server-side executable (`taxsimlocal`), which we include in our bundle. However, it only works on Windows. If you're on a MAC or a linux machine, you need to install `taxsim`'s client from its website [[link](#)].

Once you have ensured `taxsim` is running, run **robustness_tax.do**. This will produce the results (data, not figures) of the post-tax series.

Step 4: Edit & Run **counterfactuals.do**

Edit the global macro `rawdataroot` as described in step #1. Edit the global macros `absmob_indir`, `absmob_sourcedir`, and `absmob_resultsdir` as described in step #2.

Run **counterfactuals.do**. This will generate the baseline and robustness counterfactual results by calling `absmob.ado` repeatedly with different counterfactual income distributions. It will output these results into a "Counterfactuals" folder within `absmob_resultsdir`.

Step 5: Edit & Run **make_bounds.m**

Edit the variable `absmob_resultsdir` on line 5 of `make_bounds.m` to the same `absmob_resultsdir` as in step #2. Run `make_bounds.m`, making sure that both **make_bounds.m** and **maxminv100.m** are in the same folder.

Make_bounds.m calculates the bounds on baseline absolute mobility under all possible copulas. Results are saved as `.txt` files in a new folder named `matlab Matlab` in `absmob_resultsdir`.

Step 6: Edit & Run **draw_figs.do** and **draw_app_figs.do**

Edit the global macros `absmob_indir`, `absmob_sourcedir`, and `absmob_resultsdir` as described in step #2 for both **draw_figs.do** and **draw_app_figs.do**. The

Run both **draw_figs.do** and **draw_app_figs.do** to generate the main text figures and the appendix figures, respectively. Figures will be output to the `Baseline` folder of `absmob_results.dir`. The format of the figures is controlled by the global macro `image_suffix`, which is set to `png` by default.