**Replication materials for paper "Are marriage-related taxes and Social Security benefits holding back female labor supply? " by Margherita Borella, Mariacristina De Nardi, Fang Yang, edited in The Review of Economic Studies.**

Most results in the paper require the solution and simulation of the life-cycle model using a C code and the manipulation of survey data using Stata.

The numerical results are produced in two steps. First, we use c programs to solve the model. Then we use Matlab to produce figures and results in tables. The c programs were last run on a **32-core Intel server with 1024 GB of RAM, 12 TB of fast local storage**. The c program uses parallel programing and requires system of Microsoft Windows 10, 32 core of computer CPU, and the required software is Microsoft visual Studio Community 2017 and Microsoft .NET Framework Version 4.8.04084. In addition, it requires OpenMP compiler. The Matlab programs run in version R2018a.

We first describe the steps to solve and simulate the dynamic problem and to estimate its structural parameters. Then we describe how to prepare the survey data used.

The file FigureTableList.xlsx contains a complete list of all Figures and Tables and the program used to obtain them.

**I. Setup the solution, simulation and estimation of the life-cycle problem**

The following set of folders contain all the files needed to setup the C project that solves, simulates and estimates the life-cycle problem:

1. Folder \1945\_revisedRESTud is for 1945 cohort

\InputFullModel is copied from \Input\InputFullModel. Files aM.txt lifetimeY.txt in this folder are from \case4

1. Folder \1945\_revisedRESTud\_case5 is for 1945 cohort, Appendix L. Correlated earnings and life expectancy

\InputFullModel is the same as \1945\_revisedRESTud\InputFullModel

1. Folder \1955\_revisedRESTud is for 1955 cohort

\InputFullModel is copied from \Input\InputFullModel55. Files aM.txt lifetimeY.txt in this folder are from \case4

* **To replicate all the results for the 1945 and 1955 cohort, please read the file steps.docx, which contains the step-by-step procedure to run all the c codes and reproduce what is in the paper.**
* **To replicate figures and tables, run PlotAll.m in Matlab. See FigureTableList.xls for detailed listed of programs used to produce each table and figure.**

**II. Survey data**

All inputs for the dynamic model are based on two publicly available datasets: the Panel Study of Income Dynamics (PSID) and the Health and Retirement Survey (HRS).

1. **PSID**

The PSID is available without cost to researchers and analysts ([link](https://psidonline.isr.umich.edu/)). Please register to be able to download and use the data set.

To construct our main data set, we used data from the main Family data files from 1968 to 2013. To replicate our analysis, please download all the Family files from 1969 to 2013 and the Cross-Year Individual file (the most recent one will be ok), and save them into the folder INPUT\_MARRIAGE\_REPLICA\PSID\DATA.

The Family data files are then automatically unzipped and saved in STATA format using the PSID tools called by our STATA do files.

For the PSID wealth files, the STATA PSID tools cannot be used. Please download the wealth files for the waves 1984 1989 1994 1999 2001 2003 2005 2007 and unzip them into the folder “.\PSID\DATA\STATA\PSIDwlth\_raw”. Again, our do-files will then save the data into STATA format.

The datasets used in estimating the inputs and the moments for the C program are constructed by running “..\PSID\DO\_FILES\RUN\_ALL.DO”. Before running the run\_all.do file, edit the files run\_all.do and set\_up.do to set your chosen locations for the survey data, model data, do files and output files.

The estimation of the tax functions is done in a folder separate from the main PSID analysis. The data set we are using to compute the tax functions is taxes\_readytoest\_2020.dta, saved in the same folder as the do files (\INPUT\_MARRIAGE\_REPLICA\PSID\DofilesTaxsim)

1. **HRS**

The University of Michigan Health and Retirement Study (HRS) is publicly available for users registered to its site. Please register to the official site before using our replication material. The data set used is constructed using both the RAND longitudinal file and the Exit interviews, and is saved in the file \INPUT\_MARRIAGE\_REPLICA\HRS\dataready.dta

1. **Folder Structure**

The INPUT folder contains the programs needed to prepare the input and to draw figures and tables in the paper and in the appendix. It uses STATA version 14 and a user-written exe file to estimate the wage shock process. This exe file is compiled from a Fortran 90 using the IMSL libraries. The original fortran source is also available.

Sub-Folders:

1. PSID: contains PSID data and programs
2. HRS: contains HRS data and programs
3. InputFullModel and InputFullModel55: contain the input files. These are copied into the relevant folders to be used as inputs for the C code.

HOW TO REPLICATE THE INPUT FILES:

1. Run the RUN\_ALL.DO Stata do-file in the \PSID\DO\_FILES subfolder. This step prepares the PSID part of the input for the c code, as well as prepares material for tables and figures in the text.
2. Run the RUN\_ALL.DO Stata do-file in the \HRS subfolder. This step prepares the HRS part of the input, as well as prepares material for tables and figures in the text.