Description of Data and Programs

Xiaodong Fan
Monash University

Ananth Seshadri
University of Wisconsin-Madison

Christopher Taber
University of Wisconsin-Madison

May 9, 2023

1 Introduction

This document describes the data and programs used in "Estimation of a Life-Cycle Model with Human Capital, Labor Supply, and Retirement," by Xiaodong Fan, Ananth Seshadri, and Christopher Taber.

There are two main folders.

Folder "SIPP_data" contains the Survey of Income and Program Participation (SIPP) data and Stata do files to generate data moments as well as the data generating processes (DGP).

Folder "BP_MPI" contains the Fortran program to solve and estimate all variations of the life-cycle models on the paper. The program is written in Fortran with MPI and should be run at a cluster which supports Fortran and MPI.

All Stata programs are compatible with Stata 16 or newer.

2 Folder "SIPP_data"

Please unzip hclsdat.zip to obtain hclsdat.dta, which is the original SIPP data, in the same location.

The do_usedat.do is the Stata do file which cleans the data and generates data moments as well as the data generating processes (DGP). This do file utilizes the following two files as well:

i) dta_cpi.dta: the CPI data.

ii) prog_usa_tax.do: the Stata do file of USA federal income tax code in 2004, under "head of household."

3 Folder "BP_MPI"

This folder contains the Fortran program which solves and estimates all variations of the life-cycle models on the paper. The program is written in Fortran with MPI and should be run at a cluster which supports Fortran and MPI.

3.1 Subfolder "bpfcodes_com"

3.1.1 f90 files

The following seven f90 files are part of the main source codes (alphabetically ordered)

- cmaxee.f90: optimization subroutines.
- comfuncs.f90: common functions.
- debugprint.f90: mainly for debugging.
- distance.f90: simulation and calculating distance between simulated moments and data moments.
- globvar.f90: define global variables.
- gmmse.f90: calculating standard errors of parameter estimates.
- init.f90: initialization, including defining variables, reading in initial data files.

3.1.2 Subfolder "data"

These data files are used in the program as the data generating process (DGP) or moments.

i) raw_A_max.raw: the asset upper bound.
ii) raw_AIME_range.raw: the range of AIME estimated from the CPS data. The AIME space in the program is a function of this range.
iii) raw_aimeshare.raw: the $share_{min}(t)$ from the SIPP data for age 52 to 80.
iv) raw_Cvarcov.raw: the variance-covariance matrix estimated from the data moments.
v) raw_H_range.raw: the range of human capital estimated from the SIPP data. The H space in the program is a function of this range.
vi) raw_health_CPS_hs.raw: health (excellent, good, bad, or disabled) distribution and transition, the employment rates of individuals with different health status, and their differences (excellent vs good, good vs bad, bad vs disable).

vii) raw_moments_consumption.raw: the adult equivalent consumption data from the Consumer Expenditure Survey.
viii) raw_sipp_ageobs.raw: first and last ages of each individual is observed in the SIPP data.
ix) raw_sipp_moments_age_v2.raw: data moments from the SIPP.
x) raw_sipp_moments_age_workpt.raw: part-time employment rate from the SIPP.
xi) raw_sipp_moments_ss.raw: Social Security application moments from the SIPP.
xii) raw_sipp_moments_unemp.raw: wage depreciation after unemployment spell, from the SIPP.
xiii) raw_sipp_Mt.raw: family status and transitions, from the SIPP.
xiv) raw_sipp_spinc.raw: distribution of spousal income, from the SIPP.

3.1.3 Subfolder "Fcodes"

Commonly used subroutines.

3.1.4 Subfolder "Lbfgsb"

The optimization routine "setulb" used in the program.

3.1.5 Subfolder "nr"

Commonly used subroutines from the Fortran numerical recipes.

3.2 Subfolder "main"

3.2.1 f90 file

- The main.f90 is the main program code which calls all other subroutines.
- The "Makefile" is for compiling the program on Linux.

3.2.2 Subfolder "Output"

- txt_parin9.txt: initial parameter values. It has the parameter estimates on Table 3 of the baseline model.
- pars_mapping.csv: list of parameters to be estimated.
- All other files, including Stata do files: please refer to readme_dofiles.pdf for detailed descriptions.