Capital and Labor:
The Factor Income Composition of Top Incomes in the United States, 1962-2006

Anthony B. Atkinson and Christoph Lakner[*]

Readme for the replication code

Content

1. Introduction
2. Overview of .do files
3. Names of key variables in input files
4. Overview of figures
5. Overview of tables
6. Overview of specifications

1. Introduction

This readme file accompanies the replication code for Atkinson and Lakner "Capital and Labor: The Factor Income Composition of Top Incomes in the United States, 1962-2006".[1]

Directories can be set by running the first few lines of code in the master .do file ("0 AL master 1.do"). Users should create the following folders within their directory:

1. "do files": save .do files here
2. "generated data": save .dta files here
3. "graphs": save .png and .pdf files here
4. "tables": save .txt files here

The raw data used in this paper are proprietary and cannot be posted online. We use the Statistics of Income public use files (PUF) by the US Internal Revenue Service (IRS) over the period 1962 to 2006, which are (nearly) annual public-use samples based on federal income tax returns. These data that have been used in several papers by Thomas Piketty, Emmanuel Saez and Gabriel Zucman (e.g. Piketty and Saez, 2007; Saez and Zucman, 2016; Piketty et al., 2018). The complete set of files and their documentation are available through the NBER. More information on these data is available at http://users.nber.org/~taxsim/gdb/. The replication files contain a .do file which creates the key variables used in the analysis from the raw data, largely based on the .do files that are available on Gabriel Zucman's website.[2]

[*] Corresponding author: clakner@worldbank.org. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

[1] The estimations were carried out in Stata 16.1 (version dated 15 Dec 2020), running on Windows 10, 64-bit.

[2] "1a AL gen small data 1.do". This code is taken from the .do files "build60_08(SZ).do" (Saez and Zucman, 2016) and "build_small.do" (Piketty et al., 2018), available from Gabriel Zucman's website (accessed 25 September 2017).

2. Overview of .do files

Name of .do file	Name of input file	Name of output file	Description
0 AL master 1			Master file: defines directory paths and calls .do files.
1a AL gen small data 1	PUFs	ALsmall62.dta,…, ALsmall2006.dta	Prepare small files with key variables from public use files (PUF). The construction of the key variables is taken from the code used in Saez and Zucman (2016) and Piketty et al. (2018). In particular the .do files "build60_08(SZ).do" (Saez and Zucman, 2016) and "build_small.do" (Piketty et al., 2018), available from Gabriel Zucman's website (accessed 25 September 2017).
1b AL gen aux data 1	Online (Saez, Stantcheva and Zucman websites)	ALcpi.dta; ALpsz 1.dta; ALps 1.dta; ALpss 1.dta	Prepare auxiliary files (CPI, number of non-filers, total retained earnings, total corporate income tax, top share estimates from existing literature). Links to raw data included in .do file.
2a AL factor decomp 1	ALsmall62.dta,…, ALsmall2006.dta	2 AL factor decomp 2way 1.dta	Main factor income decomposition: splitting total income into two sources (capital and labor).
2b AL factor decomp 3way 1	ALsmall62.dta,…, ALsmall2006.dta	2 AL factor decomp 3way 1.dta	Appendix robustness check: factor income decomposition splitting total income into three sources (business, capital and labor).
2c AL factor decomp fig 1	2 AL factor decomp 2way 1.dta; 2 AL factor decomp 3way 1.dta	Fig. 2; Fig. A1; Fig. A2; Fig. A.3; Fig. A.7	Create figures for the factor income decomposition.
3a assoc matrices 1	ALsmall62.dta,…, ALsmall2006.dta	3 AL cond prob 1.dta; 3 AL assoc matrices 1.txt	Main estimation of the association matrices. Two output files: (a) "3 AL cond prob 1.dta": Input file for the figures showing the conditional probabilities. (b) "3 AL assoc matrices 1.txt": Tables of survival association matrices and their differences included in the paper and Appendix.
3b AL assoc matrices fig 1	3 AL cond prob 1.dta; ALpss 1.dta	Fig. 3; Fig. A.4; Fig. A.5; Fig. A.8	Figures showing the conditional probabilities.
4a AL sum stats 1	ALsmall62.dta,…, ALsmall2006.dta; 2 AL factor decomp 2way 1.dta; ALcpi.dta	4a sum stats 1.dta	Estimate table of summary statistics (Table 1 in the main text).
4b AL intro inccomp 1	ALsmall62.dta,…, ALsmall2006.dta	4b inccomp BASE 1.dta; Fig. 1	Create introduction figure on income composition (Figure 1 in the main text).
4c AL appendix lit comp 1	2 AL factor decomp 2way 1.dta; ALps 1.dta; ALpsz 1.dta	Fig. A.6	Create appendix figure which compares with earlier literature.

3. Names of key variables in input files (ALsmall62.dta,…, ALsmall2006.dta)

Variables are created in "1a AL gen small data 1.do", which includes the details on the variable construction. Variables are used in the following .do files: "2a AL factor decomp 1.do"; "2b AL factor decomp 3way 1.do"; "3a assoc matrices 1.do"; "4a AL sum stats 1.do"; "4b AL intro inccomp 1.do".

Variable name	Description
dweght	Sampling weight (scaled up by 100,000)
agi	Adjusted Gross Income (AGI)
waginc	Wages excluding 401(k) contributions
peninc	IRAs, pensions and annuities
divinc	Dividends
intinc	Taxable interest income
rentinc	Net rental income
estinc	Estates and trust income
rylinc	Royalties
schcinc	Net Schedule-C income
scorinc	Net S-corporation income
partinc	Net partnership income
kginc	Blown-up capital gains

4. Overview of figures

Number	File name	Name of .do file
1	4 intro inccomp 1	4b AL intro inccomp 1.do
2	2 combined BASE_99_1	2c AL factor decomp fig 1.do
3	3 top 1 cond BASE 1	3b AL assoc matrices fig 1.do
A1	2 combined BASE_95_1	2c AL factor decomp fig 1.do
A2	2 combined BASE_99_5_1	2c AL factor decomp fig 1.do
A3	2 appendix 3way comb 1	2c AL factor decomp fig 1.do
A4	3 top 5 cond BASE 1	3b AL assoc matrices fig 1.do
A5	3 top 05 cond BASE 1	3b AL assoc matrices fig 1.do
A6	4 appendix lit comp 1	4c AL appendix lit comp 1.do
A7	2 appendix robust comb 1	2c AL factor decomp fig 1.do
A8	3 appendix robust comb 1	3b AL assoc matrices fig 1.do

5. Overview of tables

The replication code files include an Excel file, which includes the tables in the paper and the Appendix.

Number	File name	Name of .do file
1	4a sum stats 1.dta	4a AL sum stats 1.do
2-4, A1, A5, A6, A8-A13	3 AL assoc matrices 1.txt	3a AL assoc matrices 1.do
A2-A4	Excel: Based on Table 2	
A7	Text table created in Excel file.	

6. Overview of specifications

The table below gives an overview of the different specifications that are used in the main text and the Appendix (also see Table A7 in the Appendix). The first column gives the short name that is used in the code and in file names.

Short name	Description	Definition of		Negative obs.	Non-filers
		Labor	Capital		
BASE	Baseline	W+2/3*C+3/4*(P+S)	K+1/3*C+1/4*(P+S)	Dropped	Dropped
NEG	Negatives as zero	W+2/3*C+3/4*(P+S)	K+1/3*C+1/4*(P+S)	Set to zero	Dropped
KG	Capital gains	W+2/3*C+3/4*(P+S)	K+1/3*C+1/4*(P+S)+Kgains	Dropped	Dropped
NF	Non-filers	W+2/3*C+3/4*(P+S)	K+1/3*C+1/4*(P+S)	Dropped	Included
RE	Distribute RE	W+2/3*C+3/4*(P+S)	K+1/3*C+1/4*(P+S)+RE	Dropped	Dropped
RECIT	Distribute RE + CIT	W+2/3*C+3/4*(P+S)	K+1/3*C+1/4*(P+S)+RE+CIT	Dropped	Dropped
OLD	Alt. split of business	W+2/3*(C+P)	K+1/3*(C+P)+S	Dropped	Dropped

Notes: W=wages+pensions; C=Schedule-C; P=partnership; S=S-corporation; K=dividends+interest+rents+estate income+royalties; RE=corporate retained earnings; CIT=corporate income tax

The following table explains the specifications used in the three-way split of total income into business, capital and labor ("2b AL factor decomp 3way 1.do"), which are included as a robustness check in the Appendix. In all cases, the baseline income definition is used.

Short name	Description
THREE	Splitting total income into business, capital and labor
TEXS	Three-way split excluding S-corporation income (i.e. business income includes Schedule-C and partnership income)
TEXSP	Three-way split excluding S-corporation and partnership income (i.e. business income only includes Schedule-C income)

References

Piketty, T. and E. Saez, Emmanuel: 2007, 'How Progressive is the U.S. Federal Tax System? A Historical and International Perspective', *Journal of Economic Perspectives* **21**(1), 3-24.

Piketty T., E. Saez, and G. Zucman: 2018, 'Distributional National Accounts: Methods and Estimates for the United States', *The Quarterly Journal of Economics* **133**(2), 553-609.

Saez, E. and G. Zucman: 2016, 'Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data', *The Quarterly Journal of Economics* **131**(2), 519-578.