

ReadMe, 2011_281_AEJ_Pol manuscript

03-07-2013

Description of data and programs for "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities"

Thomas Piketty, Paris School of Economics
Emmanuel Saez, UC Berkeley and NBER
Stefanie Stantcheva, MIT

Corresponding author for data and programs questions is Stefanie Stantcheva, stefanie@mit.edu

Part A: US long-term analysis (Folder "US analysis")

1. usdata.do generates Figure 1 and performs the regressions from table 1, using the data file usdata.dta
Figures are exported in the files Fig1A.wmf or .eps and Fig1B.wmf or .eps.
Regression results are outsheeted by hand in Excel.
2. usdata.dta contains all the data for the US analysis since 1913 (Figure 1, Table 1).

Part B: International macro analysis (Folder "International analysis")

1. international.do generates Figures 2, 3 and 4 and performs the regressions from table 2 using the data file AllCountries_update_pre75.dta. The Figures are exported in the files fig2A.wmf, fig2A.eps, fig2B.wmf, fig2B.eps, fig3A.wmf, fig3A.eps, fig3B.wmf, fig3B.eps, fig4A.wmf, fig4A.eps, fig4B.wmf, fig4B.eps (note that numbering is off by 1 relative to paper, fig2A in programs is fig1A in paper, etc.)
Regression results are exported in table2raw.xls and formatted in Excel.
Bootstrap results are exported in the file growth_percentiles.xls.
2. AllCountries_update_pre75.dta contains all the data for the international cross-country analysis.
3. AllCountries_update_pre75.xls contains all the data for the international cross-country analysis in excel format.

Part C: US CEO compensation analysis (Folder "US CEO analysis")

This part uses commercial data, which is not freely available, but is accessible partly through the WRDS, and partly from other researchers.

Three datasets are needed which we cannot provide:

- Compustat for North America from 1969 to 2012.
- Execucomp for North America from 1992 to 2011
(both of these are available through the WRDS)
- Forbes compensation data, available from Kevin Murphy at USC

1. Dataset_creation.do uses the three raw datasets above to construct a working data set, then used in the subsequent analysis. The dataset outputted is called working.dta.

2. NIPA_allwages_clean.dta contains NIPA data on wages per sector, which is used by the do file BM_wages.

3. top_brackets.dta contains the top tax bracket for the US, used by the files BM_tables and BM_tables_800.

4. US_CPI.dta contains the US CPI series, used to deflate nominal variables and used by the files BM_tables, BM_tables800 and BM_wages.

5. UStaxes.dta contains the top Income Marginal tax rate in the US.

6. BM_tables.do performs all the regressions reported in Table 3 (except for columns 3 and 6, see below). Regression results are outsheeted/outputted in individual .tex or .xml files and manually inputted into Excel.

7. BM_tables_800.do performs a robustness check, in which only those companies from the Execucomp sample for post-1992 are used, which match the Forbes800 pre-1991 definition. These results are not reported in the paper but are provided for completion.

8. BM_wages.do performs the analysis on workers' wages reported in columns 3 and 6 of Table 3. Regression results are outputted into several different files and then manually inputted into Excel to generate Table 3.

Part D: International CEO compensation analysis (Folder "International CEO analysis")

This part of the analysis also uses confidential data, compiled by Nuno Fernandes, Miguel A. Ferreira, Pedro Matos, and Kevin J. Murphy for their paper "Are US CEOs Paid More? New International Evidence" (Review of Financial Studies, Volume 26, Number 2, 6 February 2013 , pp. 323-367(45)) Interested readers need to contact these authors for the data set. We name the dataset `data_compensation_pub.dta` for use in the do files.

1. The file `explore_ceo_graphs.do` uses `data_compensation_pub.dta` to generate the graphs in Figure 5. The outputs are the files `ceocomp_ceocontrols.wmf` (.eps, .gph) and `ceocomp_ltotalpay.wmf` (.eps, .gph).

2. `explore_tables.do` uses `data_compensation_pub` to generate the regression results for Table 4. The results are outputted in the intermediate file `table_ceocomp.tex` (.xls) before being manually formatted in Excel.

3. `taxes.dta` contains top Income Marginal Tax Rates for the relevant countries in 1996. This was compiled for the international macro analysis in part B (Description of the construction of this data is in the Appendix of the paper).