

(last update: 13-04-2017) Read Me – PYZ 2017 Distributional Series

This directory includes all files related to the series on the distribution of income and wealth in China 1978-2015 (PYZ 2017). These files are organized as follows:

PYZ2017FinalDistributionSeries.zip. This file includes the final series by g-percentiles for the WID.world distributions of pre-tax national income, private wealth and fiscal income (as well as additional non-WID.world series on distributions of non-fiscal income and raw survey income, before any correction), together with a codebook.

PYZ2017DistributionalData.xlsx. This file includes raw distributional data, including household survey tabulations and income tax tabulations (as well as reference series on aggregate household survey income and fiscal income, income tax schedules and income tax revenues).

Directory Gpinter. This directory includes the simulation results using the “Gpinter” (Generalized Pareto interpolation) web interface (<http://wid.world/gpinter>). The Gpinter interface became available in December 2016. In the present paper it was used solely for the wealth distribution simulations and the non-fiscal income adjustment (copula simulations for $y_f + y_{nf}$, with y_f = fiscal income and y_{nf} = non-fiscal income). See explanations below. Previous simulations using survey income and income tax data were done in September-November 2016 on the basis of MatLab computer codes (these codes deliver similar distributional series by g-percentile as the Gpinter interface, but they are more complicated to use; for future uses we recommend to use Gpinter).

Directory Stata Files. This directory include the final distributional series and codebook, together with all Stata-format codes and files that we use to produce these homogenous g-percentile income and wealth series out of GPinter and MatLab simulations. The do-files are in the subdirectory StataFiles/dofiles.

Directory Xls Files. This directory includes summary tables and figures in xls format using the Stata format final distributional series.

Directory Tabulations. This directory includes the income and wealth tabulations used for the MatLab simulations. The raw tabulations come directly from household surveys. The corrected tabulations include a correction based upon high-income fiscal data.

DirectoryParetoCurves. This directory includes the Pareto curves and other figures constructed using the final distributional series. The computation of upgrade factors to be applied to raw survey data on the basis of high-income fiscal data is presented in the file ChinaUrban_yfraw.xlsx and these upgrade factors are used in dofile do5_gperc_China_urban_yfupgrade.do in order to generate corrected tabulations out of raw tabulations.

Directory SimulationOneDimMatlab. This directory includes the MatLab codes to transform tabulations into g-percentiles using one-dimensional generalized Pareto interpolation techniques.

Directory SimulationTwoDimMatlab. This directory includes the MatLab codes to transform tabulations into g-percentiles using two-dimensional generalized Pareto interpolation techniques (additive techniques and copula techniques).

How to use Gpinter directory and simulations:

The Gpinter directory is organized as follows.

(1) The **subdirectory WealthTabulations** contains **BenchmarkWealthTabulations.xlsx** (=raw wealth tabulations combining wealth survey and Hurun billionaire ranking data) (=input file to be used in apps.wid.world/gpinter) and **BenchmarkWealthSeries.xlsx** (=output file coming directly out of apps.wid.world/gpinter). We also include in the subsubdirectory Variants a number of alternative wealth tabulations (based upon alternative assumptions in order to combine wealth survey and Hurun billionaire ranking data).

(2) The **subdirectory Copulas** contains **CopulasTabulations.xlsx** (as well as CopulasTabulationsUrban.xlsx and CopulasTabulationsRural.xlsx) (=raw tabulations of non-fiscal income ynf and fiscal income yf, exported by do11a_gperc_China_w.do on the basis of BenchmarkWealthSeries.xlsx) (=input files to be used in apps.wid.world/gpinter with “add up component” option) and **CopulasSeries.xlsx** (as well as CopulasSeriesUrban.xlsx and CopulasSeriesRural.xlsx) (=output files coming directly out of apps.wid.world/gpinter). We also include in the subsubdirectory Variants a number of alternative output files corresponding to other Gumbel copula parameters ($\theta=2, 2.5, 3.5, 4$ instead of $\theta=3$).

(3) **The different routines must be run in the following order. First**, run Stata do-file do17_gperc_China_adjustaverages.do in order to update the series for average income and wealth and modify all the distributional series accordingly (be careful: the do-file do17_....do extracts averages series from xlsx files rawdata_bysheet_China_wealth, rawdata_bysheet_urbanChina_raw et rawdata_bysheet_ruralChina_raw, which themselves refer to PYZ2017DistributionalData.xlsx; one must ensure that these files are up-to-date).

Next, run Gpinter with input BenchmarkWealthTabulations.xlsx and save the output into BenchmarkWealthSeries.xlsx. **Next**, run Stata do-file do11a_gperc_China_w.do in order to extract distribution series for wealth and non-fiscal income and export CopulasTabulations.xlsx (and CopulationsTabulationsUrban.xlsx, ...Rural.xlsx). **Next**, run Gpinter with input CopulasTabulations.xlsx and save the output into CopulasSeries.xlsx (do the same separately for ...Urban and ...Rural). **Next**, run State do-file do11b_gperc_China_y.do in order to extract distribution series for ynf+yf. Finally, run again do-file do17_gperc_China_adjustaverages.do in order to adjust all averages (in particular y is larger than ynf+yf).

Note on Matlab simulations. The MatLab codes and Stata do-files (other than the do-files do11a..., do11b... and do17... referred to above) were run in September-November 2016 and in principle do not need to be run again (the Gpinter interface is available since December 2016 and is easier to use). In case one wants to use these computer codes in order to reproduce the corresponding results, the various Matlab computer codes and State do-files files need to be run in the right order.

E.g. one first needs to apply SimulationOneDimMatlab/gen_gperc_bysheet.m to SimulationOneDimMatlab/rawdata/rawdata_bysheet.xls (after replacing rawdata_bysheet.xls by Tabulations/rawdata_bysheet_UrbanChina_raw.xls = raw urban survey tabulations). Matlab output appears in SimulationOneDimMatlab/output/bysheet/

Then one needs to run StataFiles/do1_gperc_China_urban_raw.do. This uses previous output in order to produce files StataFiles gperc_yf_UrbanChina_raw.dta and StataFiles/tables/table1uraw.dta (xls) and table2uraw.dta (xls), which are then used in ChinaParetoCurves_urban_raw.xls in order to compute upgrade factors on the basis of high-income fiscal data. And so on. (be careful: stata xls output to be re-saved in 1995 xls before applying matlab)