

(last update: 13-04-2017)

Codebook (list of variables) of distributional Stata files for China

The file averages_China.dta contains the following 23 aggregates variables (income and wealth variables are in current yuans, and population variables are in millions inhabitants) over 38 years (1978-2015):

year : year

adultpop : adult population

y : per adult pre-tax national income

w : per adult net private wealth

yf : per adult pre-tax fiscal income

ys : per adult survey income

ynf : per adult pre-tax non-fiscal income (private component of undistributed profits and other income flows not included in fiscal income)

ypri0 : per adult pre-tax enlarged fiscal income (=yf+ynf) (fiscal income +non-fiscal income)

ypri : per adult post-tax disposable income (=ypri0 – income taxes + cash transfers)

ypub0 : per adult indirect taxes + government capital income (including public component of undistributed profits)

ypub : per adult public spending (= ypub0 + income taxes – cash transfers) (incl. gov. surplus)

(y=ypri+ypub=ypri0+ypub0)

it : per adult income taxes

ct : per adult cash transfers

adultpopu : urban adult population

yu : per adult urban pretax national income

wu : per adult urban net private wealth

yfu : per adult urban fiscal income

ysu : per adult urban survey income

adultpopr : rural adult population

yr : per adult rural pretax national income

wr : per adult rural net private wealth

yfr : per adult rural fiscal income

ysr : per adult rural survey income

There are 18 gperc files corresponding to the equal-split-adults distribution of w , y , yf (=final WID series on the distributions of private wealth w , pre-tax national income y and fiscal income yf), and also yf_raw , ys_raw , ynf (=non-WID series on the distributions of raw fiscal income yf_raw , raw survey income ys_raw , and non-fiscal income ynf),¹ for China, urban China and rural China. I.e. $6 \times 3 = 18$ gperc files.

All gperc files contain 4826 lines (38x127 year-percentiles) and 7 variables:²

year : income year (year takes 38 values: year = 1978-2015)

p : generalized percentile (g-percentile) (i.e. proportion of population with income below the given income threshold) (multiplied by 100000) (p takes 127 values: $p=0, 1000, 2000, \dots, 99000, 99100, \dots, 99900, 99910, \dots, 99990, 99991, \dots, 99999$)

f : frequency (fraction of population with income between two consecutive thresholds) (multiplied by 100000) (i.e. $f=p[_{n+1}]-p$) (in effect f takes 4 values: $f = 1000, 100, 10, 1$) (f can be used in order to weight individual observations)

ythr : income or wealth threshold

yint : average intermediate income or wealth

ytop : average top income or wealth (i.e. average above threshold)

b : inverted Pareto coefficient (i.e. ratio between average above threshold and threshold)

¹ The distribution of yf_raw and ys_raw are identical (they are both based upon raw household survey tabulations), the only difference being the average value of yf_raw is equal to average fiscal income and the average value of ys_raw is equal to average survey income (aggregate survey income is typically around 80% of aggregate fiscal income, as estimated using national accounts series; see PYZ2017DistributionalData.xlsx).

² Note that gperc files for wealth w only cover the 1995-2015 period (i.e. the gperc for 1978-1994 are by construction equal to the 1995 gperc, with average wealth adjustment).