World Income Distribution (WYD) database: variable definitions

The creation and key features of this database are explained in detail in papers listed on this Website, and in particular B. Milanovic, True world income distribution 1988 and 1993: first calculations based on household surveys alone, World Bank Policy Research Working Paper No. 2244 (the first paper based on these data). This note provides definitions for the variables included in the accompanying file `wyd_88_05_for_release.dta`.

Definitions

contcod = World Bank country acronym

country = Country name

year = Benchmark year. There are five benchmark years: 1988, 1993, 1998, 2002 and 2005. Income or consumption data from surveys not conducted in the benchmark years are adjusted by simply assuming an unchanged distribution and deflating/inflating incomes by country's Consumer Price Index between the actual survey year and the benchmark year. Table below shows the distribution of surveys by years. For more detail, see B. Milanovic "Global inequality recalculated and updated: the effect of new PPP estimates on global inequality and 2005 estimates", *Journal of Economic Inequality*, volume 10, issue 1, 2012, pp. 1-18, and in particular Table 2.

Proximity of actual survey years to the benchmark year (in % of all surveys)

Benchmark year	1988	1993	1998	2002	2005
Survey year same as benchmark year	37	33	43	34	55
Within a year of benchmark year	71	66	70	66	85
Within two years of benchmark year	90	88	85	94	93

region = Five regions of the world as defined in this database

Dexpend = Dummy variable indicating whether the survey aggregate is consumption- or income-based (if consumption==1)

whole = Dummy variable indicating that the data refer to the whole (entire) country as opposed to referring to urban or rural areas only. The latter applies to China, India, Indonesia and Bangladesh.

group = Ordered income fractile. The poorest fractile takes the value of 1, the second poorest 2 etc.. With decile data, the highest group would be 10.

Dinc = Average per capita income of a given fractile. Expressed in domestic currency units, annualized amount. Note that Dinc may be either income or consumption per capita (depending on whether Dexpend is 0 or 1). If there are ten equally-sized fractiles, then Dinc will give average per capita income/consumption per decile.

pop = population (in million) in each fractile. For example, if for the US, we have decile data, then population in each fractile would be around 30 million (about 300 million divided by 10). Population data are obtained from the average national population in benchmark years.

Note that (1) ranking is always by household per capita income or expenditures; (2) that all members of a household are assigned the same (average) income or consumption and (3) that population is composed of individuals (not households). Thus, the poorest decile would consist of 10 percent of poorest people in a country ranked by their household per capita income (or consumption).

<u>A couple of examples</u>

Thus, `list   Dinc pop Dexpend if contcod=="IND" & year==1998 & group==3`

will give

`Dinc = 3410.873, pop =97.09, Dexpend==1`

Indicating that the third Indian fractile in benchmark year 1998 had an average per capita consumption of 3410.873 rupees per year.

To find out the entire distribution for India in benchmark year 1998, we type

`list  group Dinc pop   if    contcod=="IND"  &    year==1998`

and obtain

```
      group        Dinc      pop

      -------------------------

      1    2240.484    97.09

      2    2860.102    97.09

      3    3410.873    97.09

      4    3923.397    97.09

      5    4448.669    97.09

      6     5012.19    97.09

      7    5676.856    97.09

      8    6556.561    97.09

      9    7973.437    97.09

      10    21180.9    97.09
```