

# List of Aggregate Variables

How do Hours Worked Vary with Income?  
Cross-Country Evidence and Implications

Alexander Bick, Arizona State University

Nicola Fuchs-Schündeln, Goethe University Frankfurt, CFS and CEPR

David Lagakos, UCSD and NBER

August 17, 2017

## D List of Aggregate Variables

For those interested in using our data, there are three relevant data sets. The data can be found in the folder `Data/Cleaned_Data` of the replication files on the AER webpage. In this document we describe the list of variables included and naming conventions used in the data sets.

### D.1 `repl_master_stats.dta`

This data set contains the aggregate statistics (on the country level), which allow replication of all information (statistics, figures, tables) in the paper with the exception of those reported in Section 4.4, Section 6 and Online Appendix Section A.6.

1. `country`, `countryname`, year of the survey (`refyear`), and GDP per capita from PWT 9.0 (expenditure side real GDP; in levels [`gdppc_e9ry`] and logs [`ln_gdppc_e9ry`]), country income group (low, medium, high; `q_e9ry`), and thresholds to classify countries into the income groups (`t1_e9ry`, `t2_e9ry`).
2. `coreDummy` takes the value 1 for core countries (see Section 3.2).
3. `var_g_age(x)`
  - (a) `var` = labor input
    - i. `hwp`: hours worked per adult
    - ii. `er`: employment rate (to be precise the respective employment to population ratio)
    - iii. `hwe`: hours per worker.
  - (b) `g` = gender: `a`=all, `f`=female, `m`=male
  - (c) `a` = age
    - i. `all`: 15+
    - ii. `prime`: 25-54
    - iii. `old`: 55+
    - iv. `non_young`: 25+
    - v. 5-year age bins, only for hours worked per adult, and neither by education nor sector
  - (d) `x`= education (`edu`) or sector (`ind`)
    - i. `edu1`= less than secondary, `edu2`=secondary completed, `edu3`=more than secondary, `edu5`=non-missing education
    - ii. `ind1`=agriculture, `ind2`=manufacturing, `ind3`=services, `ind5`=non-mising sector; only hours worked per employed are decomposed by industry.
    - iii. Categories 1 to 3 are mutually exclusive, category 5 combines 1 to 3.
  - (e) `hwp_a_age_edu(n)`: Hours worked per adult, by age $\times$ education combinations
    - i. `age`: 10 year bins
    - ii. `education`: 1, 2, 3 (see above)
  - (f) Shares (`sh_`) in paid/self-employment, missing status

- i. sh\_wj\_emp: Share of paid employees over all working individuals.
  - ii. sh\_wjse\_emp: Share of paid employees plus self-employed over all working individuals
  - iii. sh\_wj\_ws: Share of paid employees over all individuals with non-missing status
  - iv. sh\_mwj\_ws\_hr\_all, sh\_mwjse\_ws\_hr\_all: Share missing wage rates over all observations with available status (for Table A.3).
- (g) others
- i. gdppc\_o9ry, ln\_gdppc\_o9ry: Output-side real GDP variables from PWT.
  - ii. robustDummy: 0 (core plus partial years), 2 (all remaining countries), see Table (1).
  - iii. n: variable for the sorting of countries by ln(GDPPC\_e9ry).
  - iv. N\_Adults: Absolute number of 15-99 year old individuals (after data harmonization and cleaning).
  - v. year of the survey (year), same as refyear; redundant but used in some Do-files to facilitate merging with external data sets.

## D.2 timeuse\_stats.dta

This data set includes all information (set of variables and countries) used in Section 4.4. These variables are available for the age group all (15-99) only.

1. country, refyear, country income group (low, medium, high; q\_e9ry)
2. p\_var\_g\_all\_m: Fraction of individuals who participated in home service production activity.
  - (a) var = home service production activities
    - i. cooking, cleaning, childcare, shopping, coll\_wf (collecting water and firewood).
  - (b) g = gender: a=all, f=female, m=male.
3. var\_g\_all\_x
  - (a) var = home production variables
    - i. cooking, cleaning, childcare, shopping, coll\_wf (collecting water and firewood).
    - ii. total\_hrs: sum of hours spent in activities.
  - (b) g = gender: a=all, f=female, m=male.
  - (c) \_x = \_m or \_uc:
    - i. \_m: Mean conditional hours spent in activity.
    - ii. \_uc: Mean unconditional hours spent in activity:  $p\_var\_g\_all\_m \times var\_g\_all\_m$ .

## D.3 decile\_hours\_wages.dta

This data set includes hours and hourly wages by deciles, separately for individuals in paid employment (Figure 5) and for all individuals (Figure C.6).

1. country, year, country income group (low, medium, high; q\_e9ry)

2. wage decile: note that composition of individuals making up each decile differs between the paid employment and paid + self-employed samples.
3. var\_x:
  - (a) var = weekly hours by decile (hours), hourly wages by decile (wage)
  - (b) x = paid employed (paid\_empl), paid + self-employed (all)