

---

# **Data Sources Documentation**

***Release 1.0***

**Lutz Hendricks**

January 13, 2016



<b>1</b>	<b>The Penn World Tables</b>	<b>3</b>
1.1	Background . . . . .	3
1.2	Obtaining the data . . . . .	3
<b>2</b>	<b>PWT: Matlab Code</b>	<b>5</b>
2.1	Directory structure: . . . . .	5
2.2	Programs . . . . .	5
2.3	Basic Steps . . . . .	5
2.4	Exercises . . . . .	6
<b>3</b>	<b>World Development Indicators</b>	<b>7</b>
3.1	Getting the Data . . . . .	7
3.2	Getting the Data Into Matlab . . . . .	7
3.3	Matlab Code . . . . .	7
3.4	Exercises . . . . .	8
<b>4</b>	<b>Barro-Lee Schooling Data</b>	<b>9</b>
4.1	Getting the Data . . . . .	9
4.2	Importing Into Matlab . . . . .	9
<b>5</b>	<b>Indices and tables</b>	<b>11</b>



Contents:



---

## The Penn World Tables

---

The PWT's main purpose is to construct panel data for GDP and its components in **constant international prices**.

### 1.1 Background

National Accounts report GDP and its components in local currency units (LCU).

We could use exchange rates to make these figures into dollars.

The data would then imply that people in low income countries are shockingly poor, not just very poor.

This would be misleading because **prices are systematically lower** in low income countries.

To compare living standards across countries (and over time), it is necessary to deflate GDP with the local price of a **consistent bundle** of goods.

#### 1.1.1 How Is This Done?

The ICP collects data on local prices every several years (**benchmark years**).

Between benchmark years, prices are interpolated.

Even though there is no theoretically “best” price index, there is a substantial theory of how to construct price indices.

The PWT does this and reports **GDP in international prices**.

This comes in 2 flavors:

- *RGDPE*: to compare living standards (think: deflated by the consumer price index)
- *RGDPO*: to compare productive capacity (think: deflated by a producer price index)

### 1.2 Obtaining the data

The data can be downloaded [here](#).

I recommend downloading the **Stata file** and converting into a **matlab dataset** using **Stat/Transfer**.

Format of the Stata file:

- each row is a country / year combination
- countries are identified by their **ISO codes**. These are 3 letter abbreviations.

- each column is a variable (except for the first 4 columns which are country / year info)

Since Mathworks plans to phase out datasets, it would make sense to then convert the whole thing into a Matlab table using `dataset2table`.

It would now make sense to make each variable into an array indexed by [country, year].

- Fortunately, someone has already done this, so you can download the result from the PWT web site.
  - Unfortunately, that file contains errors (as of 2014-Dec). So you cannot use it!
-



---

## PWT: Matlab Code

---

### 2.1 Directory structure:

baseDir:

- outDir: figures and tables
- matDir: generated mat files
- dataDir: original data files
- progDir: program files

### 2.2 Programs

The code is entirely general purpose (not specific to the course).

- go\_pwt8 startup; add dir to path
- const\_pwt8 set constants
- run\_all\_pwt8 runs everything in sequence
- import\_pwt8 imports stata file into matlab
- var\_load\_yc\_pwt8 loads one variable by [Country,year]
- country\_list\_pwt8 makes list of countries and years

### 2.3 Basic Steps

1. Download Stata file.
2. Make Stata file into a matlab dataset using [Stat/Transfer](#)
3. import\_pwt8: break the Stata file into individual variables and save them as matlab matrices, indexed by [year, country]
4. var\_load\_yc\_pwt8: loads one variable for a given set of years and countries

## 2.4 Exercises

1. Write code that imports the stata file (`import_pwt8`)
2. Write `var_load_yc_pwt8`
3. Plot the density of real output per worker in 2000 (`rgdp_density_growth821`).
4. Plot the price of consumption against real gdp per worker for the year 2000.

What do you find?

---

## World Development Indicators

---

A collection of cross country data on

- gdp and its components
- employment
- demographics
- education
- government finances
- and much more

### 3.1 Getting the Data

One typically downloads one variable at a time from [here](#).

The best format is probably xls.

One gets a file with years as columns.

### 3.2 Getting the Data Into Matlab

Replace the column headers for the years with something that makes a valid variable name (e.g. x1970).

Run through Stat/Transfer to obtain a Matlab dataset.

Load the file into matlab. Convert the data portion of the matrix into a matrix by [year, country].

Country indicators are World Bank WITS codes (essentially the same as ISO codes).

The code now directly reads the xls file

### 3.3 Matlab Code

#### 3.3.1 Directories

- prog: programs
- excel: raw xls files and generated mat files (one for each variable)

- outDir: tables and figures

### 3.3.2 Programs

- `const_wdi2013`: set constants
- `var_load_yc_wdi2013`: load one variable by [country, year]
- `country_list_wdi2013`: make a list of all countries in the data (WITS codes)

## 3.4 Exercises

1. Construct GDP per worker (i.e. per employed person) in constant international dollars.
    - (a) Save it as a Matlab matrix by [country, year].
    - (b) How does it compare with PWT data?
  2. Plot the fraction of self-employed workers against GDP per capita in constant international dollars. For the year 2000. What do you find?
  3. Download a variable from the WDI web site and import it into Matlab.
-

---

## Barro-Lee Schooling Data

---

The standard dataset for working with cross-country schooling data.

Covers educational attainment by [year, country, age group, sex].

How is this constructed?

- mainly from household surveys
- since there are not many surveys for most countries, there is a lot of interpolation

### 4.1 Getting the Data

Download Stata files from the [Barro-Lee web site](#).

Each row is a country / year.

Each column is a variable.

Each file is one sex.

### 4.2 Importing Into Matlab

Make the Stata file into a Matlab `dataset` using `Stat/Transfer`.

Write a function `var_load_yc_bl2013` that extracts one variable for a given set of years and countries.



---

## Indices and tables

---

- `genindex`
- `search`