

Technical Guide to PWT8.0

The construction of PWT8.0 has entailed a great many choices on data and methods and we have documented and explained these choices in the various documents on the PWT website (www.ggdc.net/pwt). However, it is certainly possible to argue for alternative choices and we provide the current guide, the data files and the programs to allow users to generate alternative versions of PWT8.0. This allows, for example, for more extensive types of robustness analysis, such as whether a set of econometric findings is robust to PWT's choice of a particular index number method for aggregating PPPs.

This guide consists of two parts, namely an express guide and a basic guide. In the express guide, we outline what you need to know to generate your own alternative versions using the standard parameter options. The basic guide gives a more in-depth overview of the structure of the database and an overview of the different data files and Stata programs that would allow for modifications at a more basic level.

The do-files have been written in and run on Stata 12.1. They may also run on earlier versions, but no guarantees (or support) is given for that. Also note that the Stata data format was last changed with Stata 11, so earlier versions will not be able to open the data files.

Express guide

For the most relevant alternative options, only the main do-file has to be edited, `gen_pwt.do` and then only the five parameters listed at the start of the file (lines 3-7). Running the `gen_pwt.do` program (make sure that Stata's active directory is the one with the programs) will then generate an alternative PWT dataset based on your chosen parameter set. Table 1 lists the five parameters, what they do, and the valid and default values for each of them.

For a more detailed understanding of the different parameters, please refer to the main documents related to PWT8.0, namely *The Next Generation of the Penn World Table* (Feenstra, Inklaar and Timmer 2013a; FIT1) and *PWT8.0 – A User Guide* (Feenstra, Inklaar and Timmer, 2013b; FIT2), both available through www.ggdc.net/pwt. FIT1 discusses the index number methods used in PWT8.0, placing these in the perspective of the broader literature on how to aggregate detailed PPPs into more aggregate PPPs. That discussion is helpful in the context of parameters `index1` and `index2`. FIT1 also discusses the use of multiple PPP benchmarks (parameter `bm`) and the normalization of PPPs to be equal to the US GDP deflator (parameter `norm`). FIT2 discusses how and why the PPP and National Accounts data for China have been modified from the official series, and is thus relevant for parameter `chn`. For those interested in mimicking earlier versions of PWT, the `bm` parameter is most relevant, as earlier PWT versions constructed a full time series of real GDP based on a single benchmark.¹

¹ This is not an exact replication as in PWT7 and before, the ICP PPP benchmarks for a given year would be augmented with estimates based on earlier benchmark years as well as estimates for countries that did not participate in any benchmark comparison up to that point. In addition,

Table 1, Parameters in PWT8.0, gen_pwt . do

Parameter	Function	Valid values*
index1	Chooses the index number procedure for aggregating basic heading PPPs into PPPs for household consumption, investment and government consumption**	"geks" "gk"
index2	Chooses the index number procedure for aggregating PPPs for main expenditure categories into GDP PPPs***	"geks" "gk"
bm	If set to . , all PPP benchmarks are used. To only use PPPs from a single benchmark comparison, set this to the preferred year	. 2005 1996 1985 1980 1975 1970
chn	Use either the modified PPPs and growth rates, by setting to "pwt" or use the official PPPs and growth rates, by setting to "icp"	"pwt" "icp"
norm	If set to 0, all price levels are normalised so that the values for the USA are equal to the USA GDP deflator. If set to 1, the USA=1 in all years	0 1

Notes:

* The default value used in constructing PWT8.0 is shown in bold

** The method for aggregating export and import PPPs cannot be modified

*** Choosing "geks" for index2 reduces the set of variables in the final dataset as some variables lose their meaning. This is because GK is an additive index number method (see Balk, 2008), so that shares in real GDP are sensible. In contrast, GEKS is not an additive method.

In addition to changing these five parameters, it is also straightforward to generate PWT datasets based on earlier versions of National Accounts (NA) data. As discussed in FIT2, basic NA data are prone to frequent and sometimes large revisions, which can change comparative growth performance and income rankings. Earlier NA data can be downloaded from www.ggd.net/pwt; to use these in the PWT calculations, rename the chosen file (e.g. na_data_v2009.dta) to na_data.dta and replace the existing na_data.dta in the Data folder with this alternative file.

Basic Guide

With the provided data files and programs, it is also possible to modify the computation of PWT at a more basic level. This should only be undertaken by expert users since changing one part of the programs can have unexpected consequences in other parts. Changing the five parameters listed in Table 1 or changing the NA data should always lead to output that is comparable to the standard PWT database, but (obviously) this cannot be guaranteed when changing other elements.

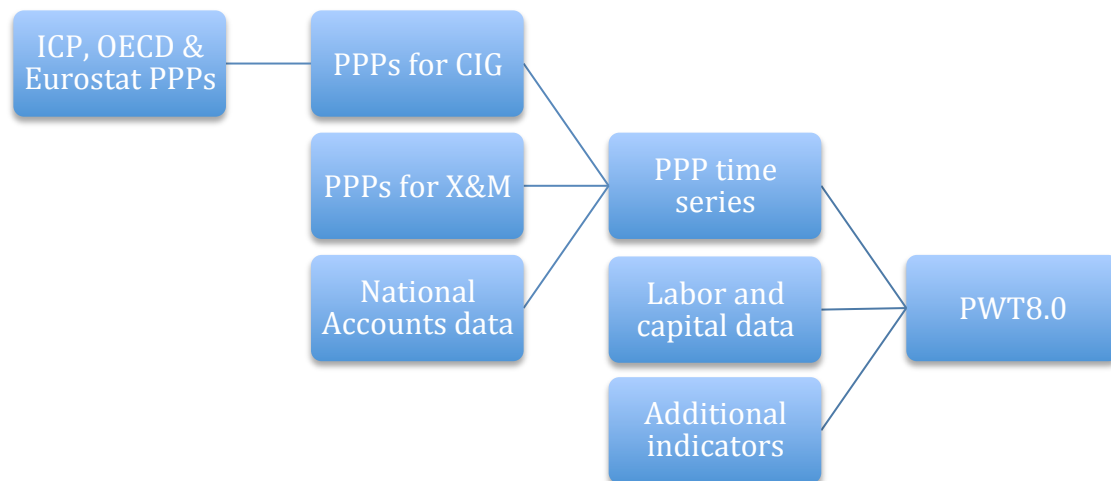
This Basic Guide serves as a roadmap to the construction of PWT8.0, discussing the data files, programs and how they fit together. At a more detailed level, the programs include frequent comments and explanations to clarify what is done with various bits of code.

there have been methodological changes over time relating to index number methods and, in PWT8.0, the inclusion of export and import PPPs.

Schematic overview

The schematic below provides an overview of how the different data sources are combined in construction PWT8.0. The remainder of this guide discusses the data files relevant for the different parts of this schematic and the programs that move the data from one part to the other.

Figure 1, Schematic overview of the construction of PWT8.0



Throughout, PPPs are normalized to relative price levels by dividing the PPP (in local currency unit per U.S. dollar) by the exchange rate (in local currency unit per U.S. dollar). We use PPPs and relative price levels interchangeably here. The acronym CIG stands for (household) Consumption, Investment and Government (consumption) and X&M stands for Exports and imports, with $GDP=C+I+G+X-M$.

All data files are organized by the variables countrycode and year. Countrycode refers to the ISO (alpha-3) country identifier, see for the full correspondence of codes to country names: e.g. http://en.wikipedia.org/wiki/ISO_3166-1_alpha-3. In contrast with earlier versions of PWT, we strictly adhere to the formal list. The two differences are Romania, where we use ROU rather than ROM and the Democratic Republic of Congo, where we use COD rather than ZAR.

Support programs

Files: index.do eks_func_create.do fischer_func_create.do
geks_func_create.do gk_func_create.do laspeyres_func_create.do
paasche_func_create.do

A number of programs are included to support calculations. The program index.do transforms a series of growth rates into an index with a given base value. The programs (eks, fischer, geks, gk, laspeyres, paasche) all refer to index number methods and are used for aggregating detailed PPPs into more aggregate PPPs. These files contain Mata functions, that are called on in Mata sections of the Stata do-files. See e.g. Balk (2008) for a more detailed discussion of each of these methods.

ICP, OECD & Eurostat prices

Files: ICPxxxx.xls, where xxxx is 1970, 1975, 1980, 1985 and 1996

For confidentiality reasons, we cannot provide the prices at basic heading level for the 2005 ICP comparison; for the 1995-2010 Eurostat comparisons; or the 1996, 1999, 2002, 2005 and 2008 OECD comparisons. Researchers with a project that requires data at this detailed level can contact the respective organizations:

- World Bank: icp@worldbank.org
- Eurostat: estat-ppp@ec.europa.eu
- OECD: ppp.contact@oecd.org

PPPs for CIG

Files: gen_bm_level.do level_analysis.do ppp_icp.dta ppp_es.dta
ppp_oecd.dta ppp_icp_CHN_ICPbh.dta

gen_bm_level.do imports the ICP basic heading level PPPs from Excel and then uses level_analysis.do to compute aggregate PPPs for household consumption, investment, and government consumption (CIG). The two index number methods given under index1 (GEKS and GK) are applied in this program and in the next stage, the desired method is used for further analysis. Since the basic heading level data for ICP 2005 cannot be made publicly available, the gen_bm_level.do program does not generate a new version of the ppp_icp.dta data (which is used in the next step). Likewise, the CIG PPPs based on the Eurostat benchmarks (ppp_es.dta) and on the OECD benchmarks (ppp_oecd.dta) is the most detailed that can be made publicly available. In each of the three ppp_***.dta files, pl_c_geks is the relative price level of household consumption, computed using the GEKS index number procedure and normalized so that the USA=1 in each year. The variable pl_c_gk is the relative price level of household consumption, computed using the GK index number procedure, where the GDP price level of the USA=1 (so pl_c_gk for the USA differs from 1). The file ppp_icp_CHN_ICPbh.dta is the same as ppp_icp.dta, except that the Chinese data are calculated using the ICP basic heading PPPs, rather than the adjusted basic heading PPPs, see FIT2 and Deaton and Heston (2010). This file is used if parameter chn is set to "icp".

PPPs for X&M

Files: xm_bec_data_q.dta

This file contains PPPs for exports and imports by 1-digit broad economic category (BEC), see <http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=10>. These have been constructed as described in Feenstra and Romalis (2012). Only the first 6 BEC categories are included as the BEC-7 is a small group with miscellaneous products and a highly variable composition across countries. Also given are the exports and imports by BEC. The export PPP for BEC-1 (Food and beverages) is pl_x1 and exports of BEC-1 products is given by x1.

National Accounts Data

Files: na_data.dta tot_xm.dta us_trade_prices.dta basic_ch2_data.dta

The main National Accounts data file, `na_data.dta`, is described in more detail in *National Accounts in PWT8.0*. All the data are in local currency units. The variables starting with `v_` are in current prices and the variables starting with `q_` are in constant prices. Population and exchange rate data is also included; for more information on the distinction between market exchange rates and (market+estimated) exchange rates, see the document *Exchange Rates in PWT8.0*.

Also included under this category is the file `tot_xm.dta`. This file provides total merchandise exports (`tot_x`) and total merchandise imports (`tot_m`), used in the next step for combining export and import PPPs with National Accounts data. At that point, we assume that the exports (imports) for the 6 BEC categories in `xm_bec_data_q.dta` are representative for all merchandise exports (imports).

The next data file is `us_trade_prices.dta`, which provides the US export and import prices by BEC category. These are used in the next step to correct for US inflation in combination with the export and import PPPs. These price indices are taken from the US Bureau of Economic Analysis (BEA), specifically the table 'Price Indexes for Exports and Imports of Goods and Services by Type of Product', which covers the period since 1967. Before 1967, the prices for total goods exports and total goods imports are used. Export prices for fuels and lubricants are taken from the BLS for the period since 1983 and assumed equal to the trend in fuel and lubricant import prices before 1983.

The final data file is `basic_ch2_data.dta`, which provides the alternative National Accounts data for China, constructed by Harry Wu of Hitotsubashi University. How these data are used and the precise sources are given in the *National Accounts in PWT8.0* document. The implications are discussed in FIT2.

PPP time series

Files: `gen_oecd_es_ts.do` `gen_ppp_ts.do`

At this point, the PPP data are combined with NA data to estimate PPPs for the main expenditure categories (CIGX&M) for every year. This involves interpolating between benchmark years, using NA price trends information, and extrapolation before the first and after the last benchmark year using NA price trends, see FIT1 and FIT2 for more information on this approach.

In the first program, `gen_oecd_es_ts.do`, the (non-EU) OECD and Eurostat CIG PPP data are combined and a panel of PPPs for the period 1995-2010 is estimated. That panel is then inserted in the program `gen_ppp_ts.do` and combined with ICP benchmark data. In technical terms, this implies that we impose 'block fixity' on the CIG PPPs with the US as the only connection between the OECD/Eurostat block and the other countries in the world. Similarly, the (non-EU) OECD PPPs are combined with the Eurostat PPPs through Germany. See World Bank (2013) for more details on what this type of linking implies.

Export and import PPPs are also inserted in `gen_ppp_ts.do`, and merchandise exports and imports are combined with NA data, see *National Accounts in PWT8.0* for more details on this combination. With all the benchmark information on CIGX&M PPPs and the NA data, we then estimate a full panel of PPPs through interpolation and extrapolation. Then GDP PPPs are estimated for

every year using the index number procedure given in `index2`. This provides all the PPP data that is necessary in the final compilation of the PWT variables.

Labor and capital data

Files: `pwt_emp_data.dta` `capital_results.dta` `lab_share_data.dta`
`rkna_data.dta` `BL2013_MF_v1.3.dta` `process_barrolee.do` `bl_data.dta`

These are the data used in computing comparative levels of total factor productivity (TFP) and TFP growth. The sources and methods in constructing these data are described in detail in the document *Capital, labor and TFP in PWT8.0*. The data items in the data files are used almost one-for-one in PWT8.0. The exceptions are `BL2013_MF_v1.3.dta`, which is the data on the years of schooling taken directly from Barro and Lee (2012, www.barrolee.com); the program `process_barrolee.do`, which extracts the relevant information on the average years of schooling of the population aged 15 and up; and `bl_data.dta`, which contains the output of `process_barrolee.do`.

Additional indicators

Files: `esh_cor_data.dta` `statcap_data.dta` `names_units.dta`

The data items in these files are also used one-for-one in PWT8.0. The data in the first two files, `esh_cor_data.dta` and `statcap_data.dta`, are discussed in more detail in FIT2. The file `names_units.dta` contains the country names and the names of the local currency unit.

PWT8.0

Files: `gen_pwt.do` `gen_final_pwt.do`

As noted in the Express Guide, the file `gen_pwt.do` is the main file that calls on all the relevant do-files in the earlier stages and combines the data into the final form. This involves computing the different real GDP versions and computing TFP levels and TFP growth. This file also labels the variables to indicate how the PPPs are computed for each country and each year, which exchange rate is used, and whether an observation is considered an outlier; see the document *Outliers in PWT8.0*. The final call is on `gen_final_pwt.do`, which organizes and labels all the variables for PWT8.0.

References

- Barro, Robert J. and Jong-Wha Lee (2012), "A new data set of educational attainment in the world, 1950-2010" *Journal of Development Economics*, forthcoming.
- Deaton, Angus and Alan Heston (2010), "Understanding PPPs and PPP-based National Accounts" *American Economic Journal: Macroeconomics* 2(4): 1-35.
- Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2013a), "The Next Generation of the Penn World Table", mimeo, available at: www.ggd.net/pwt.
- Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2013b) "PWT8.0: a User's Guide", mimeo, available at: www.ggd.net/pwt.
- Balk, Bert M. (2008), *Price and Quantity Index Numbers: Models for Measuring Aggregate Change and Difference*, Cambridge University Press, Cambridge.
- World Bank (ed. 2013) *Measuring the Real Size of the World Economy*, World Bank: Washington DC.