User Guide to PWT 9.0 data files

Table of Contents

PWT 9.0

This file, in Excel or Stata format, is the Penn World Table and includes the main variables. In its variables and their construction, this data file closely resembles PWT versions 8.0 and 8.1. New users are therefore advised the following:

1. Read Section I of Feenstra et al. (2015),[1] available to download through http://www.rug.nl/ggdc/productivity/pwt/related-research-papers/the_next_generation_of_the_penn_world_table.pdf. This clarifies what types of real GDP variables are available in PWT and which should be used when.
2. The remainder of Feenstra et al. (2015) provides the theoretical background and details on the empirical implementation. The online appendix, included in the document linked above, provides further details on specific elements of PWT, such as the construction of capital and labor data.
3. To gain a broader understanding of the choices that were made in constructing PWT and some of the 'health warnings', read 'PWT 8.0 – a user guide' available for download through:http://www.rug.nl//ggdc/productivity/PWT/related-research-papers/pwt_80_user_guide.pdf. The arguments and results apply broadly to PWT 9.0 as well.
4. Further details on specific parts of the data are available on the PWT 8.0 release page (http://www.rug.nl/ggdc/productivity/pwt/pwt-releases/pwt8.0), specifically details on how PWT 8.0 and above relates to PWT 7.1 (and earlier), information on the exchange rate variable and on how outliers are identified and marked in the PWT data. At the end of this document, we give an overview of changes to the 'outlier' classification, which follows the earlier method.
5. The 'What is new in PWT 9.0' document on the PWT 9.0 release page details the changes relative to PWT 8.0 and 8.1. This includes new relative price data from ICP 2011, extended

[1] Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table" *American Economic Review*, 105(10), 3150-3182.

and revised national accounts data, and improved data on labor and capital; see also below for those data files.

Experienced users are advised to start with document 5 and move backwards as needed.

National Accounts data

One of the major inputs into PWT is National Accounts (NA) data on gross domestic product (GDP) at current and constant prices,[2] in local currency units. GDP is also broken down (in current and constant prices) by major expenditure categories, household consumption, investment (and gross fixed capital formation), government consumption, exports and imports. PWT 9.0 includes a new vintage of NA data, covering the period up to 2014; see the 'What is new in PWT 9.0' document for more discussion of revisions compared to the previous data. We primarily source these data from the UN National Accounts Main Aggregates Database, see http://unstats.un.org/unsd/snaama/Introduction.asp. More discussion of these data can be found in the 'National Accounts in PWT 8.0' document, available at http://www.rug.nl/ggdc/docs/national_accounts_in_pwt80.pdf.

This document discusses how we use National Accounts from official sources for the period 1950–1969 for the Netherlands and the United States. We have since expanded this to Brazil, Sweden, Japan and Israel.

ICP benchmark data

The PPP and expenditure data at a detailed level are available in the so-called benchmark years of the International Comparison Program (ICP). We can provide these detailed figures for the ICP benchmarks of 1970, 1975, 1980, 1985 and 1996. Access to the detailed data for 2005 and 2011 can only be given by the ICP Global Office at the World Bank, see http://icp.worldbank.org for the procedure and request forms.

Trade detail

PWT includes relative prices for exports (pl_x) and imports (pl_m) and reports the share of exports (csh_x) and imports (csh_m) in real GDP at current PPPs. Underlying these figures are the relative prices and shares for exports and imports by Broad Economic Category (BEC). Those underlying figures are given in the Trade Detail data file.

Labor detail

PWT includes information on total employment (emp), average hours worked (avh), human capital (hc) and the share of labor income in GDP (labsh). The compilation of employment data involves combining information from various sources; the variable i_emp in the labor detail file indicates which observation is sourced where. The document 'Human Capital in PWT 9.0' details how we compile data on average years of schooling and how we use this information to construct a human capital index. The variable yr_sch gives the average years

[2] 'Constant prices' is a somewhat dated term in National Accounts, as the widespread adoption of chained index numbers means that data cannot be said to be in prices of a particular year. Various alternative terms are therefore used, including 'real GDP', the 'quantity of GDP' and the 'volume of GDP'. The aim is the same, namely to have a GDP series that does not reflect changes in prices.

of schooling we use and variable source gives the source we use for each country. Note that for Barro-Lee we use version 2.0 of their data; for Cohen-Soto-Leker, we use the 2014 update of the Cohen-Soto database, available at http://www.parisschoolofeconomics.eu/en/cohen-daniel/international-educational-attainment-database/.

The labor income share data is based on a set of possible alternative measures. The variable i_labsh indicates which choice we make for each country and i_labsh2 indicates whether the data in a given year are based on observed data or assumed constant from the first or last observation or interpolated between observations. In addition, all alternative labor share measures (comp_sh and lab_sh#, where $\# = 1, \ldots 4$) and included as well are whether those are observed or assumed constant or interpolated.

Capital detail

The capital stock variables in PWT (ck and rkna) and the relative price of the capital stock (pl_k) is built up from investment data by asset. At the most detailed level, we estimate investment for 9 assets but at that level of detail, the degree of estimation required for most countries is such that this would not be generally usable or informative. So instead, we provide information for 4 assets: structures (including residential and non-residential), machinery (including computers, communication equipment and other machinery), transport equipment and other assets (including software, other intellectual property products, and cultivated assets).

For each of those 4 assets, the capital detail file includes information on investment at current national prices (the Ic_* variables), the investment deflator (Ip_*), the current-cost net capital stock (Kc_*), the capital stock deflator (Kp_*) and capital consumption at current prices (Dc_*). The relationship between these variables is as follows:

Investment at constant national prices for asset a: $I_{at} = Ic_{at}/Ip_{at}$

Capital stock at constant national prices for asset a: $K_{at} = (1 - \delta_{at})K_{at-1} + I_{at}$

Current-cost net capital stock: $Kc_{at} = Kp_{at} \times K_{at}$

Depreciation rate of asset a: $\delta_{at} = Dc_{at}/Kc_{at}$

At the level of 4 assets, not all variables can derived since, for example, the depreciation rate of an asset depends on the capital stock, while the capital stock is computed based on the depreciation rate. At the level of 9 assets, the capital stock deflator equals the investment deflator and the depreciation rate are chosen exogenously and remain constant over time. The depreciation rates are as follows: residential structures 1.1%, non-residential structures 3.1%, computers 31.5%, communication equipment 11.5%, other machinery 12.6%, transport equipment 18.9%, software 31.5%, other intellectual property products 15% and cultivated assets 12.6%.

Also note that we cannot provide the (cross-country) relative prices time series at the asset detail level that underlie the pl_k variable since that would require more detailed information

than we can provide from ICP 2005 and ICP 2011. For the other benchmark years, the ICP benchmark data provide the relative prices.

Expenditure share correlations

As discussed in the 'User Guide to PWT 8.0', it becomes intrinsically harder to accurate compare prices and living standards if countries differ more in their economic structure and spending patterns. To compare each country to the United States, PWT includes the cor_exp variable, which in the ICP benchmark years gives the correlation between the vector of detailed expenditures for each country with the vector for the United States. For assessing the reliability of any bilateral comparison, the expenditure share correlations data file provides the bilateral correlations.

Outlier update

Country/year observations in which relative prices and (sometimes) relative income levels take on values that are hard to reconcile with ICP benchmark information have been labeled as outliers, starting with PWT 8.0. The approach taken is discussed in more detail there (see http://www.rug.nl/ggdc/docs/outliers_in_pwt80.pdf). With new relative price and National Accounts data as well as broader country coverage, the classification of outliers is revisited for PWT 9.0. This has led to the following *changes* to the classification:

- Data for **Brunei**, **Burundi**, **El Salvador** and **Zimbabwe** are no longer labeled as outliers for part of their sample.
- **United Arab Emirates:** this country is included in PWT 9.0, but not in 8.x because its first ICP participation was in the 2011 round. The country was included in earlier PWT versions based on alternative estimates of relative prices. In PWT 9.0, relative price data for the 1970–1973 period suddenly drops compared to the post 1973 period, to levels below what is observed in ICP benchmarks. Moreover, since it is a high-income country, with a GDP per capita approximately equal to that in USA in 2011, a very low price level seems hard to defend. Furthermore, throughout the 1970–1985 period, relative income levels of 3 or higher (USA=1) are typically observed, again far outside the range of observed relative income levels for benchmark or interpolated observations. Presumably for that reason, PWT 7.1 excluded data for the country for this period. We therefore label the data for the 1970–1985 period as outliers.
- **Turks and Caicos Islands:** for the period 1970–1973, price levels exceeding twice the US level are observed, falling outside the range observed for benchmark or interpolated observations. We therefore label the data for this period as outliers. Implausibly large swings in terms of trade also lead to outliers in 2003 and 2014 with relative price levels for GDP^o exceeding 10 (with USA=1).