**Online Appendix to:**

**The Next Generation of the Penn World Table\***

by

*Robert C. Feenstra*
University of California, Davis and NBER

*Robert Inklaar*
University of Groningen

*Marcel Timmer*
University of Groningen

Revised, April 2015

**Appendix A:** **Proof of Theorems**

*Proof of Theorem 1:*

Because the outputs $y_k$ are feasible for country $k$, but not optimal at the prices $P_j$, it

follows that $r_k(P_j, v_k) \geq P'_j y_k$ This establishes the first inequality below and the second is

established similarly: [1]

$$\frac{r_j(P_j, v_j)}{r_k(P_j, v_k)} \leq \left( \frac{P'_j y_j}{P'_j y_k} \right) \quad \text{and} \quad \frac{r_j(P_k, v_j)}{r_k(P_k, v_k)} \geq \left( \frac{P'_k y_j}{P'_k y_k} \right).$$

Using the Gerschenkron effect in (10) it follows that:

$$\frac{r_j(P_j, v_j)}{r_k(P_j, v_k)} \leq \left( \frac{P'_j y_j}{P'_j y_k} \right) < \left[ \left( \frac{P'_j y_j}{P'_j y_k} \right) \left( \frac{P'_k y_j}{P'_k y_k} \right) \right]^{1/2} < \left( \frac{P'_k y_j}{P'_k y_k} \right) \leq \frac{r_j(P_k, v_j)}{r_k(P_k, v_k)}.$$

Now define $\Pi = \lambda P_j + (1-\lambda) P_k$, and consider forming the ratio $r_j(\Pi, v_j) / r_k(\Pi, v_k)$. By

construction, this ratio equals the right-hand side of the above expression when $\lambda = 0$ and the

left-hand side when $\lambda = 1$. Therefore, there exists a value of $\lambda \in (0,1)$ at which this ratio equals

the Fisher ideal price index in the middle of the above expression. QED

*Proof of Corollary:*

This follows immediately from Theorem 1 by indexing country $j$ with the time subscript $t$, and

then treating country $k$ as identical to country $j$ in year $t$-1. QED

*Proof of Theorem 2:*

Let us write the vector of reference prices $\Pi$ as $\Pi = (\pi_1, ..., \pi_N)$ and the vector of

endowments as $v_j = (v_{1j}, ..., v_{Lj})$ for country $j$. Since the reference prices are equal in the revenue

functions $r_j(\Pi, v_j)$ and $r_k(\Pi, v_k)$, we can be quite flexible about how they appear. In particular,

suppose that $r_j(\Pi, v_j)$ is of the form:

---

[1] These inequalities are due to Malmquist (1953), as noted by Diewert (2008).

1

$$\ln r_j(\Pi, v_j) = \ln h_j(\Pi) + \sum_{l=1}^{L} \alpha_{lj} \ln v_{lj} + \sum_{l=1}^{L} \sum_{i=1}^{N} \beta_{lij} \ln v_{lj} \ln \pi_i + \frac{1}{2} \sum_{l=1}^{L} \sum_{m=1}^{L} \gamma_{lm} \ln v_{lj} \ln v_{mj}, \qquad \text{(A1)}$$

where $h_j(\Pi)$ is homogenous of degree one. The function $\ln r_k(\Pi, v_k)$ is specified similarly but with $k$ replacing $j$.

Without loss of generality we can assume that $\gamma_{lm} = \gamma_{ml}$, and notice that only these second-order parameters do not depend on countries $j$ or $k$. To interpret this restriction, note that we can multiply the endowments $v_{lj}$ in (A1) by country-specific Harrod-neutral productivity terms $\lambda_{lj}$, which after simplification are absorbed into the parameters $\alpha_{lj}$ and also into $h_j(\Pi)$, but the parameters $\gamma_{lm}$ on $\ln v_{lj} \ln v_{mj}$ remain unchanged. Very general types of country-specific sectoral productivity terms can be introduced through the parameters $\beta_{lij}$ and through the function $h_j(\Pi)$. But the condition that the second-order parameters $\gamma_{lm}$ do not differ across countries restricts the generality of factor-augmenting productivity differences between countries (beyond Harrod-neutral) that can be introduced.

Using the definition of the factor prices $w_{jl}^* = \partial r_j(\Pi, v_j) / \partial v_{jl}$ and the assumption that $r_j(\Pi, v_j)$ is homogeneous of degree one in endowments, we have $\dfrac{w_{lj}^* v_{lj}}{\sum_m w_{mj}^* v_{mj}} = \dfrac{\partial \ln r_j(\Pi, v_j)}{\partial \ln v_{jl}}.$

It follows from (A1) that:

$$\frac{w_{lj}^* v_{lj}}{\sum_m w_{mj}^* v_{mj}} = \alpha_{lj} + \sum_{i=1}^{N} \beta_{lij} \ln \pi_i + \sum_{m=1}^{L} \gamma_{lm} \ln v_{mj}. \qquad \text{(A2)}$$

Then using the definition of $A_j$ and $A_k$ along with (A1) and (A2), we can compute:

$$\ln\left(A_j A_k\right)^{1/2} = \frac{1}{2}\left[r_j(\Pi, v_j) - r_k(\Pi, v_j) + r_j(\Pi, v_k) - r_k(\Pi, v_k)\right]$$

$$= \left[r_j(\Pi, v_j) - r_k(\Pi, v_k)\right] - \frac{1}{2}\left[r_j(\Pi, v_j) + r_k(\Pi, v_j) - r_j(\Pi, v_k) - r_k(\Pi, v_k)\right]$$

$$= \left[r_j(\Pi, v_j) - r_k(\Pi, v_k)\right]$$
$$- \frac{1}{2}\left[\sum_{l=1}^{L}\left(\alpha_{lj} + \alpha_{lk} + \sum_{i=1}^{N}(\beta_{lij} + \beta_{lik})\ln\pi_i\right)\ln\left(\frac{v_{lj}}{v_{lk}}\right) + \sum_{l=1}^{L}\sum_{m=1}^{L}\gamma_{lm}\ln v_{lj}\ln v_{mj} - \sum_{l=1}^{L}\sum_{m=1}^{L}\gamma_{lm}\ln v_{lk}\ln v_{mk}\right]$$

$$= \left[r_j(\Pi, v_j) - r_k(\Pi, v_k)\right]$$
$$- \frac{1}{2}\left[\sum_{l=1}^{L}\left(\alpha_{lj} + \alpha_{lk} + \sum_{i=1}^{N}(\beta_{lij} + \beta_{lik})\ln\pi_i\right)\ln\left(\frac{v_{lj}}{v_{lk}}\right) + \sum_{l=1}^{L}\sum_{m=1}^{L}\gamma_{lm}(\ln v_{mj} + \ln v_{mk})(\ln v_{lj} - \ln v_{lk})\right]$$

$$= \left[r_j(\Pi, v_j) - r_k(\Pi, v_k)\right] - \frac{1}{2}\left[\sum_{l=1}^{L}\left(\frac{w_{lj}^* v_{lj}}{\sum_m w_{mj}^* v_{mj}} + \frac{w_{lk}^* v_{lk}}{\sum_m w_{mk}^* v_{mk}}\right)\ln\left(\frac{v_{lj}}{v_{lk}}\right)\right],$$

where the first equality follows directly from the definition of $A_j$ and $A_k$; the second equality

follows from simple algebra; the third equality follows from the translog formula in (A1); the

fourth equality follows from algebra on the double-summations that $\gamma_{lm} = \gamma_{ml}$ ; and the final

equality follows from the share formula in (A2). QED

**Appendix B: PWT Variables and Formulas**

*Detailed PPP estimation*

We start from the following three datasets:

1. Benchmark data from ICP, OECD and Eurostat, providing data for detailed consumption and

   investment products[2] on prices $p_{ijt}^k$ and quantities[3] $q_{ijt}^k$ for product $i$, country $j$, year $t$ and

   with $k$ equal to household consumption $C$, government consumption $G$ or investment. Prices

   are given relative to those in the United States and expenditures are in current national prices.

---

[2] What we refer to as 'products' are typically referred to as 'basic headings' in the benchmark data.
[3] The relative quantity of any variable is obtained by dividing its relative expenditure by its relative price.

2.  Data for Broad Economic categories (BEC) of exports and imports from Feenstra and

    Romalis (2014). Prices $p_{ijt}^k$ and quantities $k_{ijt}$ have been aggregated from 4-digit Standard

    Industrial Trade Classification (SITC) categories to six, one-digit BEC categories *i*, country *j*,

    year *t* and with *k* equal to exports *x* or imports *m*. Prices are expressed relative to the United

    States and expenditures are in current national prices.

3.  National Accounts data of expenditure and price deflators $P_{jt}^{NA,k}$. Data are only available for

    the spending categories $k = C, I, G, X, M$ in country *j* at time *t*. Expenditures are in current

    national prices and the price deflators are relative to a benchmark year with 2005 = 1.

    In the first step, we aggregate the benchmark price data as to obtain price indexes for

overall household consumption, government consumption and investment using a GEKS

procedure. For a given year *t*, define the Fisher price index for spending category *k* between

country *j* and country *h* as:

$$P_{jh}^{F,k} \equiv \left[ \left( \frac{p_j'^k q_j^k}{p_h'^k q_j^k} \right) \left( \frac{p_j'^k q_h^k}{p_h'^k q_h^k} \right) \right]^{1/2}. \tag{B1}$$

Then the GEKS price index for spending category *k* between country *j* and country *h*, computed

over a set of countries $c = 1, ..., C$ to obtain a transitive comparison, is:

$$P_{jh}^{G,k} \equiv \prod_{c=1}^{C} \left( P_{jc}^{F,k} P_{ch}^{F,k} \right)^{1/C}. \tag{B2}$$

For every benchmark comparison, we apply equation (B2) to get price indexes for *C, I* and *G*.

These are all initially scaled to with US=1, but for easier comparison of magnitudes over time,

we multiply these by the US deflator $P_{USA,t}^{NA,k}$. Denote the resulting price indexes, with USA in

2005=1, as $P_{jt}^{G,k}$. For exports and imports, we use the price indexes from Feenstra and Romalis

(2014) defined at the level of six, one-digit Broad Economic Categories (BEC) *i*, such as fuel or

consumer goods. These have also been aggregated from the 4-digit SITC data using a GEKS procedure. Multiplying these by US price deflators gives the corresponding $P_{ijt}^{G,k}$ .[4]

Since the ICP benchmark comparisons have been done in only six years – 1970, 1975, 1980, 1985, 1996 and 2005 – and not all countries participated in every comparison, we have a dataset for 167 countries[5] and comparatively few observations per country.[6] To get a complete set of price indexes, we interpolate between benchmarks and extrapolate for years before the first and after the last available benchmark.[7] This procedure is illustrated schematically in Figure C1.

For a hypothetical country that participated only in the 1996 and 2005 benchmarks, we use interpolated price indexes for the 1997-2004 period, while the years up to 1995 and from 2006 onwards are based on extrapolated price indexes. Extrapolation to year $t$–1 if price index data for year $t$ are available is done using the price deflators from the National Accounts:

$$P_{jt-1}^{G,k} = P_{jt}^{G,k} \times \frac{P_{jt-1}^{NA,k}}{P_{jt}^{NA,k}} \tag{B3}$$

Interpolation takes the benchmark price indexes as given but uses the pattern of inflation in intervening years to determine how much of the overall change in price indexes between

**Figure B1: Schematic Illustration of a Hypothetical Price Index Computation**



| Extrapolation | Interpolation | Extrapolation |
|---|---|---|

1996      2005
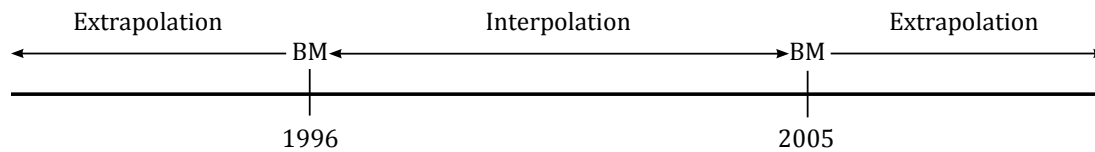
*Note*: BM stands for 'benchmark'. This hypothetical country participated only in the 1996 and 2005 benchmarks.

---

[4] These still have subscript *i* for each BEC.
[5] This is the full set of (currently existing) countries that participated in at least one ICP comparison.
[6] Data from OECD and Eurostat provide more frequent benchmark comparisons for the period after 1995.
[7] For exports and imports, data is continuous for most countries after 1984, so the process to get a complete time series primarily involves extrapolation before 1984.

benchmarks should fall in any given year. So for the hypothetical country of Figure B1, this would be:

$$P_{j,t}^{G,k} = P_{j,1996}^{G,k} \times \frac{P_{j,t}^{NA,k}}{P_{j,1996}^{NA,k}} \left( \frac{2005 - t}{2005 - 1996} \right) + P_{j,2005}^{G,k} \times \frac{P_{j,t}^{NA,k}}{P_{j,2005}^{NA,k}} \left( \frac{t - 1996}{2005 - 1996} \right). \tag{B4}$$

This approach to interpolation and extrapolation is similar in spirit to the approach of Rao *et al* (2010), who also discuss a method for estimating price indexes for a full set of years and countries using benchmark price indexes and National Accounts deflators. The key distinction is that we always force the price series to be equal to the benchmark estimates, while this is a special case of Rao *et al* (2010); see also Hill (2004).

### *Real GDP computation*

Given price indexes for the major expenditure categories and expenditures on these categories from the National Accounts, we estimate real GDP. To simplify notation compared to the previous discussion, let us denote for each country $j$ and time period $t$ the GEKS price index $p_{ijt} \equiv P_{jt}^{G,i}$, for $i = C, I, G$, while the GEKS prices indexes for export and imports are denoted by $p_{ijt}^{x}$ and $p_{ijt}^{m}$, respectively, across the $i = 1,\ldots,6$ BEC categories. We therefore have a set of 15 'products' for every country and year: three for domestic final expenditure $C, I$ and $G$, six for exports, $X$ and six for imports, $M$. Using the Geary-Khamis system, we define the following set of reference prices:

$$\begin{aligned} \pi_{it} &= \sum_{j=1}^{C} \left( p_{ijt} / PPP_{jt}^{o} \right) q_{ijt} \Big/ \sum_{j=1}^{C} q_{ijt} \,, \text{ for } i = C, I, G, \\ \pi_{it}^{x} &= \sum_{j=1}^{C} \left( p_{ijt}^{x} / PPP_{jt}^{o} \right) x_{ijt} \Big/ \sum_{j=1}^{C} x_{ijt} \,, \text{ for } i = 1,\ldots,6, \\ \pi_{it}^{m} &= \sum_{j=1}^{C} \left( p_{ijt}^{m} / PPP_{jt}^{o} \right) m_{ijt} \Big/ \sum_{j=1}^{C} m_{ijt} \,, \text{ for } i = 1,\ldots,6, \end{aligned} \tag{B5}$$

where the PPP for each country is defined using nominal $GDP_{jt}$ as:

$$PPP_{jt}^o = \frac{GDP_{jt}}{\sum_{i=C,I,G} \pi_{it} q_{ijt} + \sum_{i=1}^{6} \pi_{it}^x x_{ijt} - \sum_{i=1}^{6} \pi_{it}^x m_{ijt}} . \tag{B6}$$

The system of equations (B5) and (B6) can be solved recursively and has a unique solution up to

a normalization, as shown in Feenstra *et al* (2009). The normalization that we choose is

$PPP_{USA,t}^o = P_{USA,t}^{NA,GDP}$ , so that the PPP for the USA is equal to the GDP deflator (and 2005=1).

Whether the solution to the system is also positive is less clear *a priori*, because imports

(products 10-15) are subtracted in the numerator and denominator. Only if trade shares are not

too large is it possible to rule out negative solutions (see Feenstra *et al*, 2009).

One complication is that the 15 products for which we have prices and quantities do not

make up total GDP. Our prices of exports and imports only cover merchandise trade, but net

trade in services also contributes to total GDP.[8] Furthermore, in some countries there is a

statistical discrepancy between total expenditure $C + I + G + X - M$ and total GDP. For

example, India defines total GDP from the income side and total expenditure may be larger or

smaller than GDP. We therefore define $R_{jt}$ as residual component of GDP.

Given the set of reference prices and PPPs in every year, most of the variables in PWT

can be computed. Table B1 shows these variables and corresponding formulas in logical order,

so that previously defined variables can be used in subsequent formulas. The remaining

computed variables, relating to factor inputs and productivity, are discussed in more detail in

Appendix C.

---

[8] Note the in some countries, merchandise exports or imports are larger than National Accounts total exports or imports. In those cases, we use National Accounts trade numbers and assuming net trade in services is zero.

**Table B1: PWT Variables and Formulas**

| PWT variable | Formula |
| --- | --- |
| $CGDP_{jt}^{o}$ | $\sum_{i=C,I,G} \pi_{it}q_{ijt} + \sum_{i=1}^{6} \pi_{it}^{x}x_{ijt} - \sum_{i=1}^{6} \pi_{it}^{m}m_{ijt} + \dfrac{R_{jt}}{PPP_{jt}^{o}} \equiv \dfrac{GDP_{jt}}{PPP_{jt}^{o}}$ |
| $PL\_GDP_{jt}^{o}$ | $PPP_{jt}^{o}\big/\varepsilon_{jt}$ |
| $PL\_GDP_{jt}^{e}$ | $\left( \sum_{i=C,I,G} p_{ijt}q_{ijt} \Big/ \sum_{i=C,I,G} \pi_{it}q_{ijt} \right)\Big/\varepsilon_{jt} \equiv PPP_{jt}^{q}\big/\varepsilon_{jt}$ |
| $CCON_{jt}$ | $\sum_{i=C,G} p_{ijt}q_{ijt} \Big/ PPP_{jt}^{q}$ |
| $CDA_{jt}$ | $\sum_{i=C,I,G} p_{ijt}q_{ijt} \Big/ PPP_{jt}^{q}$ |
| $CGDP_{jt}^{e}$ | $GDP_{jt}\big/PPP_{jt}^{q}$ |
| $\dfrac{RGDP_{jt}^{o}}{RGDP_{jt-1}^{o}}$ | $\left[ \left( \dfrac{\sum_{i=C,I,G} \pi_{it-1}q_{ijt} + \sum_{i=1}^{6} \pi_{it-1}^{x}x_{ijt} - \sum_{i=1}^{6} \pi_{it-1}^{m}m_{ijt}}{\sum_{i=C,I,G} \pi_{it-1}q_{ijt-1} + \sum_{i=1}^{6} \pi_{it-1}^{x}x_{ijt-1} - \sum_{i=1}^{6} \pi_{it-1}^{m}m_{ijt-1}} \right) \times \left( \dfrac{\sum_{i=C,I,G} \pi_{it}q_{ijt} + \sum_{i=1}^{6} \pi_{it}^{x}x_{ijt} - \sum_{i=1}^{6} \pi_{it}^{m}m_{ijt}}{\sum_{i=C,I,G} \pi_{it}q_{ijt-1} + \sum_{i=1}^{6} \pi_{it}^{x}x_{ijt-1} - \sum_{i=1}^{6} \pi_{it}^{m}m_{ijt-1}} \right) \right]^{1/2}$ |
| $\dfrac{RGDP_{jt}^{e}}{RGDP_{jt-1}^{e}}$ | $\left[ \left( \dfrac{\sum_{i=C,I,G} \pi_{it-1}q_{ijt} + (X_{jt} - M_{jt})/PPP_{jt}^{q}}{\sum_{i=C,I,G} \pi_{it-1}q_{ijt-1} + (X_{jt-1} - M_{jt-1})/PPP_{jt-1}^{q}} \right) \times \left( \dfrac{\sum_{i=C,I,G} \pi_{it}q_{ijt} + (X_{jt} - M_{jt})/PPP_{jt}^{q}}{\sum_{i=C,I,G} \pi_{it}q_{ijt-1} + (X_{jt} - M_{jt})/PPP_{jt}^{q}} \right) \right]^{1/2}$ |
| $CSH\_i_{jt}$ | $\pi_{it}q_{ijt}\big/CGDP_{jt}^{o} \ , \ i=C,I,G$ |
| $CSH\_X_{jt}$ | $\sum_{i=1}^{6} \pi_{it}^{x}x_{ijt} \Big/ CGDP_{jt}^{o}$ |
| $CSH\_M_{jt}$ | $-\sum_{i=1}^{6} \pi_{it}^{m}m_{ijt} \Big/ CGDP_{jt}^{o}$ |
| $CSH\_R_{jt}$ | $\left( R_{jt}\big/PPP_{jt}^{o} \right)\Big/CGDP_{jt}^{o}$ |
| $PL\_i_{jt}$ | $\left( p_{ijt}q_{ijt}\big/\pi_{it}q_{ijt} \right)/\varepsilon_{jt} \ , \ i=C,I,G$ |
| $PL\_X_{jt}$ | $\left( \sum_{i=1}^{6} p_{ijt}^{x}x_{ijt} \Big/ \sum_{i=1}^{6} \pi_{it}^{x}x_{ijt} \right)/\varepsilon_{jt}$ |
| $PL\_M_{jt}$ | $\left( \sum_{i=1}^{6} p_{ijt}^{m}m_{ijt} \Big/ \sum_{i=1}^{6} \pi_{it}^{m}m_{ijt} \right)/\varepsilon_{jt}$ |

## Appendix C: Capital, labor and productivity

The productivity level in country $j$ relative to country $k$ was defined prior to Theorem 2:

$$\left(A_j A_k\right)^{1/2} = \left. \frac{r\left(\Pi, v_j\right)}{r\left(\Pi, v_k\right)} \right/ Q_T\left(v_j, v_k, w_j^*, w_k^*\right), \tag{C1}$$

where $Q_T\left(v_j, v_k, w_j^*, w_k^*\right)$ is the Törnqvist index of factor endowments. In this appendix we

describe the measurement of the factor endowments and factor shares. For that, it is helpful to

write equation (C1) in terms of PWT variables:

$$\frac{CTFP_{jt}}{CTFP_{USAt}} = \left. \frac{CGDP_{jt}^o}{CGDP_{USAt}^o} \right/ Q_{j,USA,t} , \tag{C2}$$

where,

$$
\begin{aligned}
Q_{j,USA,t} = &\tfrac{1}{2}\left(LABSH_{jt} + LABSH_{USAt}\right)\left(\frac{EMP_{jt}}{EMP_{USAt}} \frac{HC_{jt}}{HC_{USAt}}\right) \\
&+ \left[1 - \tfrac{1}{2}\left(LABSH_{jt} + LABSH_{USAt}\right)\right]\left(\frac{CK_{jt}}{CK_{USAt}}\right)
\end{aligned}
\tag{C3}
$$

and where $LABSH_{jt}$ denotes the share of labor in total factor income in country $j$ and year $t$,

$EMP_{jt}$ denotes employment, and $HC_{jt}$ denotes a measure of human capital based on the average

years of schooling from Barro and Lee (2012). Similarly, we define productivity growth over

time as:

$$\frac{RTFP_{jt}^{NA}}{RTFP_{jt-1}^{NA}} = \left. \frac{RGDP_{jt}^{NA}}{RGDP_{jt-1}^{NA}} \right/ Q_{jt,t-1} , \tag{C4}$$

where,

$$
\begin{aligned}
Q_{j,t,t-1} = &\tfrac{1}{2}\left(LABSH_{jt} + LABSH_{jt-1}\right)\left(\frac{EMP_{jt}}{EMP_{jt-1}} \frac{HC_{jt}}{HC_{jt-1}}\right) \\
&+ \left[1 - \tfrac{1}{2}\left(LABSH_{jt} + LABSH_{jt-1}\right)\right]\left(\frac{RK_{jt}^{NA}}{RK_{jt-1}^{NA}}\right)
\end{aligned}
\tag{C5}
$$

The remainder of this appendix describes how the variables $CK_{jt}$, $RK_{jt}^{NA}$, and $LABSH_{jt}$ are computed, drawn from Inklaar and Timmer (2013).

### *Capital stocks*

Capital stocks are estimated based on cumulating and depreciation past investments using the perpetual inventory method (PIM). This section first discusses how investment by asset is estimated. Given the long-lived nature of many assets, it is important to start the PIM from an initial capital stock and the method used to estimate these is discussed next. Finally, we show the implications of the more detailed investment data for cross-country depreciation patterns and the relative capital stock levels.

### i) *Investment at current and constant prices*

There is a wide range of assets in which firms (and governments) can invest in and these tend to have widely varying asset life spans. A common shortcut method is to ignore this heterogeneity and estimate capital input based on a common and constant assumed asset life. But this ignores important changes in investment composition over time and differences across countries. The work of Caselli and Wilson (2004), for instance, shows there are considerable differences in the composition of investment across countries: for example, richer countries tend to invest more in computers.

For PWT8, we develop a dataset with investment in up to six assets, shown in Table C1 with their geometric depreciation rates. These rates are assumed to be common across countries and constant over time. As the breakdown by asset is not readily available for all countries, we use a variety of sources in compiling the investment data.

**Table C1, Assets covered and geometric depreciation rates**

| Asset | Depreciation rate |
|---|---|
| Structures (residential and non-residential) | 2% |
| Transport equipment | 18.9% |
| Computers | 31.5% |
| Communication equipment | 11.5% |
| Software | 31.5% |
| Other machinery and assets | 12.6% |

Notes: depreciation rates are based on official BEA deprecation rates of Fraumeni (1997).

We first distinguish structures, transport equipment and machinery. We do this based on OECD National Accounts, country National Accounts, EU KLEMS ([www.euklems.org](www.euklems.org)) and ECLAC National Accounts (Economic Commission for Latin America and the Caribbean). That still leaves many countries with incomplete data, so we use the commodity-flow method (CFM) whereby investment in an asset is assumed to vary with the economy-wide supply (production + imports - exports) of that asset. This approach has also been used by Caselli and Wilson (2004), though without the constraint that investment had to add up to gross fixed capital formation in the National Accounts and without covering investment in structures. The CFM method uses data on value added in the construction industry from the UN National Accounts Main Aggregates Database; imports and exports of equipment from UN Comtrade and Feenstra's World Trade Flows database; and industrial production from UNIDO. The detailed expenditure data underlying the ICP PPP benchmarks is used to fix investment shares at different points in time and the CFM is used to measure the trends over time. In a second step, we use data from EU KLEMS and from The Conference Board on information and communication technology (ICT) investment and WITSA on ICT expenditure to split up investment in machinery into investment in computers, communication equipment, software and other machinery. This second step can

only be done for a subset of countries as ICT investment data is not available for all countries. This provides us with a dataset on investment at current national prices.

For data on investment prices over time, we use EU KLEMS, OECD National Accounts, ECLAC or UN National Accounts. This last source only provides a deflator for overall investment, which is most obviously problematic for ICT assets that have shown rapidly declining prices in countries with enough data, such as the US. For ICT assets, we therefore assume that the US price trend also applies to countries for which we have no specific data from other sources, with an adjustment made for overall inflation using the GDP deflator. For many countries, though, only the total investment deflator is used for non-ICT assets.

## ii) *Initial capital stocks*

We have very long time series of investment, back to 1950 for numerous countries, but to also provide good estimates in earlier years, we have to start from an initial capital stock. We have chosen to apply a harmonized procedure for all countries. Based on our choice of an initial capital stock, we then estimate capital stocks using the perpetual inventory method. The choice for an initial capital stock procedure is a consequential one, in particular because structures have such long asset lives, and thus low depreciation rates. With a 2% annual depreciation rate and investment data since 1950, almost 30 percent of the 1950 capital stock is still in use in 2011.[9] For countries with data since 1990, such as the former Soviet republics, the surviving fraction is almost two-thirds, so the procedure used for estimating the initial capital stock is important.

Nehru and Dhareshwar (1993) discuss a number of alternatives for estimating this initial capital stock, including production function estimates and choosing an initial capital/output ratio. Their preferred approach, originally proposed by Harberger (1978), is to use the steady-state

---

[9] Calculated as $(0.98)^{61}=0.29$.

relationship from the Solow growth model:

$$K_0 = \frac{I_0}{g + \delta} \ . \qquad \text{(C6)}$$

The initial capital stock $K_0$ for an asset is related to investment in the initial year, the (steady-state) growth rate of investment $g$ and the depreciation rate $\delta$. This requires the strong assumption that all economies were in a steady state in the first year for which data is available and that a reasonable steady-state growth rate of investment can be identified. Harberger (1978) initially chose a three-year average, while Nehru and Dhareshwar (1993) (effectively) use the average growth between 1950 and 1973 and Caselli (2005) uses the average growth until 1970 (which means a 10 to 20-year average growth rate given his selection of countries with data since at least 1960).

Nehru and Dhareshwar (1993) were dissatisfied with assuming an initial capital/output ratio, for reasons they do not spell out in detail. However, we argue that this method actually leads to superior results, in particular in early years of the sample and in transition economies, where the data is available for a limited period of time and where the early years were particularly turbulent. Under this approach, the initial capital stock is estimated as:

$$K_0 = Y_0 \times k, \qquad \text{(C7)}$$

where $Y_0$ is GDP in the initial year and $k$ the assumed capital/output ratio $K/Y$. To motivate the choice for $k$, Figure C1 plots capital/output ratios in 2005, where capital is summed over all assets, against GDP per capita. The figure includes the 142 countries with investment data since at least 1970. The figure shows considerable variation in capital/output ratios around a median value of 2.7. The least squares regression line indicates that there is no systematic relationship between GDP per capita and capital/output ratios.

**Figure C1: Capital/Output Ratio versus log GDP per Capita, in 2005**
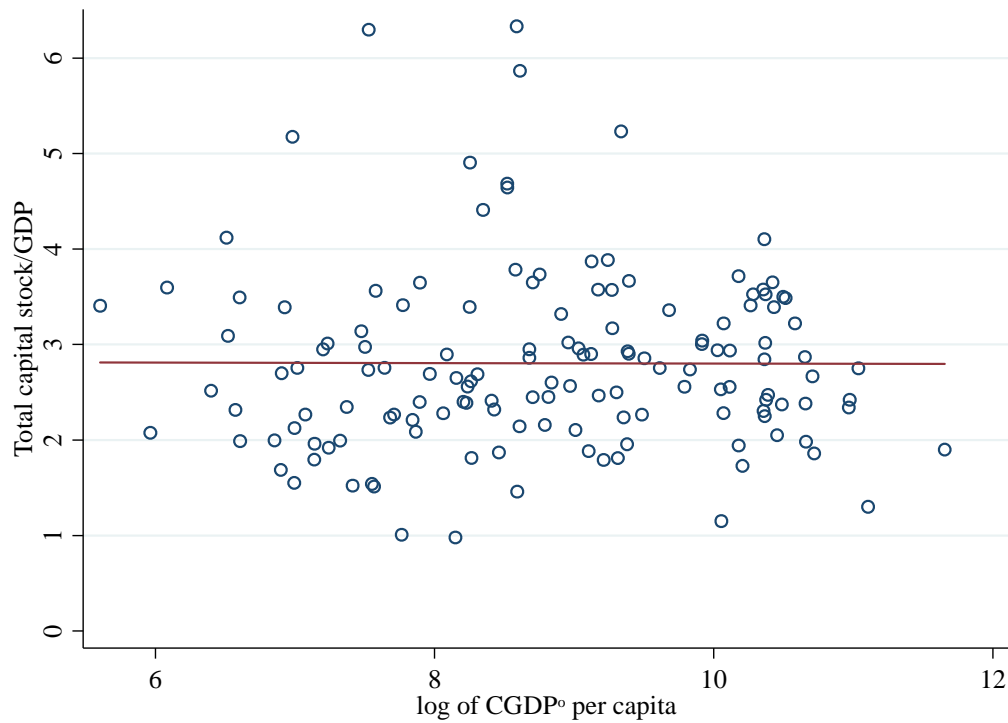


Source: PWT8.1
Note: Capital/output ratios are measured using data at current national prices. The plot excludes countries with investment data that start in a year later than 1970.

 

To solidify this finding, Table C2 shows a number of regressions, aiming to explain variation in capital/output ratios with GDP per capita. Note that the capital/output ratio is at current national prices, so does not reflect differences in inflation or relative prices of the capital stock versus the price of GDP. Column (1) shows the regression on the data in Figure C1, so with data for 2005 and including only countries with investment data since at least 1970. Column (2) includes all countries and years and shows a similarly insignificant relationship between the GDP per capita level and the capital/output ratio.

| | (1) 2005 since 1970 | (2) Total stock Full sample | (3) Structures | (4) Machinery | (5) Transport equipment | (6) Computers | (7) Communication equipment | (8) Software |
|---|---|---|---|---|---|---|---|---|
| log CGDP$^{o}$/capita | -0.00246 | 0.0399 | -0.00295 | 0.0296*** | -0.000611 | 0.00480*** | -0.0133*** | 0.00676*** |
| | (0.0524) | (0.0461) | (0.0432) | (0.00925) | (0.00475) | (0.000718) | (0.00462) | (0.000713) |
| Constant | 2.825*** | 2.554*** | 2.379*** | 0.107 | 0.145*** | -0.0296*** | 0.193*** | -0.0502*** |
| | (0.487) | (0.418) | (0.393) | (0.0801) | (0.0421) | (0.00639) | (0.0427) | (0.00627) |
| | | | | | | | | |
| Observations | 143 | 8268 | 8268 | 8268 | 8268 | 3293 | 3293 | 3293 |
| R-squared | 0.000 | 0.001 | 0.000 | 0.029 | 0.000 | 0.164 | 0.028 | 0.332 |

Note: dependent variable is the capital/output ratio, measured using data at current national prices. Robust standard errors are in parentheses; in columns (2)-(8), errors are clustered by country. Column (1) only includes data for 2005 for countries with investment data since at least 1970. The remaining columns include all countries and years in PWT8.0. The number of observations in column (6)-(8) is lower than in (2)-(5) because there is no data on investment in these assets for a range of countries. *** p<0.01, ** p<0.05, * p<0.1

Columns (3) through (8) analyze the separate components of the total capital of setting an initial capital stock, these relationships are less relevant to account for because the asset life of machinery and communication equipment is much shorter and because the use of computers and software only became widespread since the 1960s in the US and later in other countries. Furthermore, structures account for, on average, 80 percent of the value of the capital stock, so its initial stock will have the most impact on the overall results.

Table C3 shows the initial capital/output ratios (*k* in equation C7) that we assume for all countries for the non-ICT assets, based on the cross-country medians that we observe in the data over the full period. Given their short asset lives and relatively small share in total assets we set initial ICT stocks equal to zero. There is a circularity in setting initial capital stocks based on capital/output ratios that are computed based on those initial stocks. However, if we use equation (C7) to estimate initial stocks, the same median ratios are obtained.

**Table C3: Initial Capital/Output Ratios for Non-ICT assets**

| Asset | Capital/output ratio $k$ |
|---|---|
| Structures (residential and non-residential) | 2.2 |
| Transport equipment | 0.1 |
| Other machinery and assets | 0.3 |
| | |
| Total | 2.6 |

Note: initial capital/output ratios for ICT assets are set at zero.

Based on these choices, we can contrast the results based on assuming an initial capital/output ratio to results based on the more commonly-used 'steady-state' method described in equation (C6). The average growth of investment for the first ten years of the sample period is used in that equation, but the results are similar if the first five years of data are used.[10]

Table C4 shows descriptive statistics for the year 2005, comparing capital/output ratios based on assuming an initial capital/output ratio (K/Y) and based on assuming a steady-state capital level (SS). The first row compares the results for all countries and shows that the median capital/output ratio is very similar across the two approaches, but the variation is much larger using the steady-state approach. Finally, the correlation is high, but at 0.73 far from perfect.

**Table C4: Comparing Capital/Output ratios in 2005:**
**Initial ratio vs. Steady-state assumption**

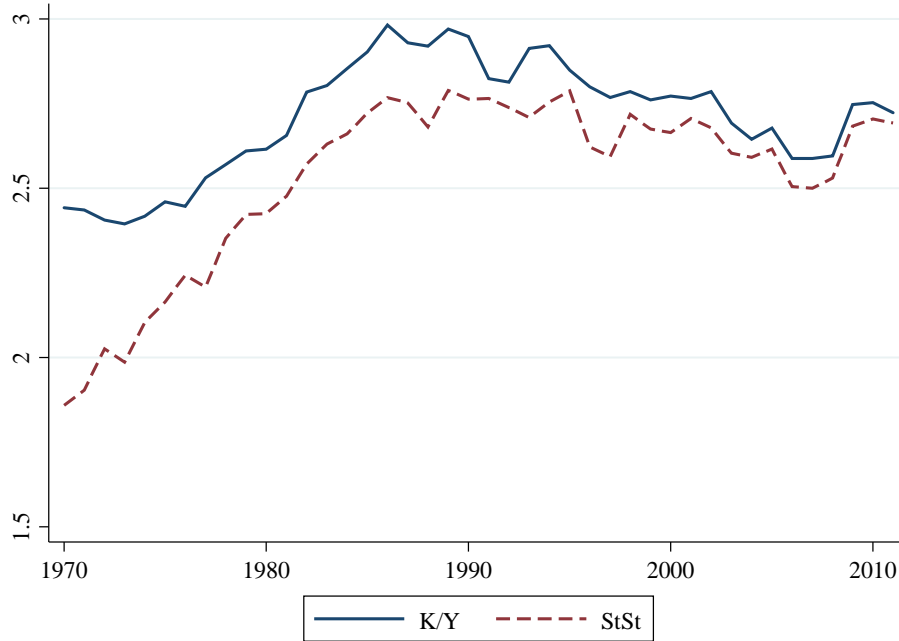| | Median | | Standard deviation | | Correlation |
|---|---|---|---|---|---|
| | K/Y | Steady State | K/Y | Steady State | (K/Y, SS) |
| All countries (167) | 2.81 | 2.74 | 1.28 | 2.10 | 0.73 |
| *Investment data starting in:* | | | | | |
| 1950-1959 (73) | 2.62 | 2.64 | 0.66 | 0.67 | 0.99 |
| 1960-1970 (69) | 2.72 | 2.63 | 1.16 | 1.23 | 0.97 |
| 1988-1990 (25) | 3.64 | 5.73 | 2.15 | 3.62 | 0.53 |

Notes: K/Y indicates that an initial capital/output ratio is assumed for the first year in which data is available; SS indicates the steady-state capital stock based on equation (C6) is used.

---

[10] The main difference is in the cross-country variation, which is much higher if only the first five years of investment data is used.

The next rows split the full sample of 167 countries by the length of the investment series. More than 70 countries have investment time series since before 1950, and another 69 countries have time series since 1970, while the final group of 25 countries has investment data for less than 25 years. In the first two groups of countries the correlation is nearly perfect and the median and standard deviation are very similar. It is in this final group that the largest differences can be seen: the steady-state approach leads to a median capital/output ratio that is much higher, variation that is much higher, and the correlation between the two approaches is quite low at 0.53. The countries in this last group are nearly all countries that emerged from the former Soviet Union, Yugoslavia and Czechoslovakia. In those newly-formed countries, the early 1990s were anything but a steady-state, involving a transition to market-based economies. Using the steady-state approach implies very high capital/output ratios, especially early in the transition period. This is because these countries saw rapidly falling GDP in the first years of their transition and thus a large increase in their capital/output ratio. To illustrate, in 1995 the steady-state approach implies a capital/output ratio in the Czech Republic of 6.5 and in Slovakia of 8.9, while in Poland and Hungary (transition countries with longer time series), the ratios are 3.4 and 3.7. In contrast, if the assumption of initial capital/output ratio is used, the 1995 capital/output ratio in the four countries varies between 3.7 and 3.9.

Figure C2 shows that differences between the initial K/Y ratio and its steady state value are much larger in earlier years. The figure shows the median capital/output ratio for the 142 countries with investment data since 1970 according to the two approaches. As the figure illustrates, from the mid-1980s onwards, the median capital/output ratio fluctuates between 2.5 and 3.1 according to both approaches, which suggests this is the standard range for the

**Figure C2: Comparing Capital/Output Ratios over Time:**
**Initial Ratio vs. Steady-State Assumption, 1970-2011**



Source: PWT8.1
Notes: K/Y indicates that an initial capital/output ratio is assumed for the first year in which data is available; StSt indicates the steady-state capital stock based on equation 0 is used. For each year, the median, standard deviation and correlation is computed for the cross-section of 142 countries with investment data since 1970 or earlier.

capital/output ratio. Assuming an initial capital/output ratio ensures that the data for these countries are continuously within this range, but applying the steady-state assumption implies that capital levels are much too low for a long period of time, starting at a median level of 1.9 and reaching 2.5 only in 1980. We therefore conclude that assuming the initial capital/output ratios from Table C3 ensures more plausible capital stocks across all countries and years.

### iii) Capital stocks and depreciation

Given an initial capital stock $K_0$, investment at constant prices $I$ and depreciation rates $\delta$, it is straightforward to compute capital stocks for asset $a$ in country $i$ at time $t$ using the Perpetual Inventory Method (PIM): $K_{ait} = (1 - \delta_a) K_{ait-1} + I_{ait}$. Multiplying this capital stock by the asset deflator $P_{ait}$ then gives capital stocks at current national prices. Compared with the typical

approach in the literature, the main benefit is that the assumption of a single depreciation rate for all countries and years is no longer necessary, since the asset composition of investment varies across countries and years.

Figure C3 illustrates the depreciation rates of the total capital stock that we get as a result of this approach.[11] The figure plots the depreciation rate in 2005 against GDP per capita and the least-squares regression line. The slope is not significantly different from zero in this year, but across all years, higher income countries tend to have higher depreciation rates. This fits with the finding from Table C2 that richer countries have higher capital/output ratios for assets with high depreciation rates: machinery, computers and software. So, for example, the U.S. had a depreciation rate in 2011 of 4.1 percent, while China had a rate of 3.1 percent. Since the capital stocks of richer countries are depreciating at a more rapid rate, the capital stock levels we estimate here will be relatively lower compared to capital stocks estimated based on a common depreciation rate across countries.

### iv) *Capital stock at constant national prices*

With capital stocks constructed for each of the assets, we construct a total capital stock at constant national prices, $RK^{NA}$. Ideally, this would be a measure of capital services, not capital stocks. A capital services measure would reflect that shorter-lived assets have a larger return in production, as indicated by the user cost of capital of each asset. As a result, buildings, which represent on average 80 percent of the capital stock at current prices, would represent a much lower share of capital services. However, the data requirements for estimating capital services are higher than for a capital stock measure. In particular, the user cost of capital of an asset

---

[11] The depreciation rate of the total capital stock is computed as $\delta_{it} = \sum_a P_{ait} \delta_a K_{ait-1} \Big/ \sum_a P_{ait} K_{ait-1}$

**Figure C3: Depreciation Rate of the Total Capital Stock and GDP per capita in 2005**
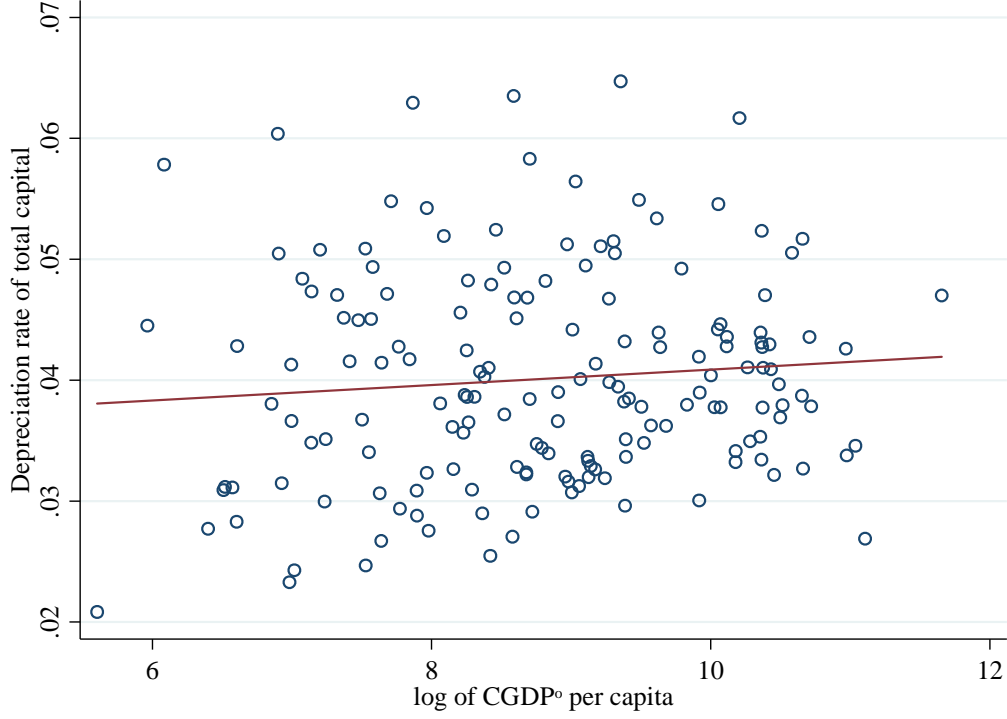


should include, alongside the depreciation rate, a required rate of return on capital and the rate of asset price inflation. Asset-specific inflation rates are not available for many countries, as mentioned above, and the required rate of return on capital is generally hard to measure well (see e.g. Inklaar, 2010). Furthermore, in countries that have experienced periods of extreme inflation in the past, any measurement error in either inflation or the rate of return would lead to substantial swings in the user cost of capital. Finally, the user cost would be needed for comparisons over time but also across countries. This implies that mismeasurement of user costs in one country would affect capital input estimates for other countries as well.

We therefore use the total capital stock as our measure of capital input. The $RK^{NA}$ variable is constructed as a Törnqvist aggregate of the individual asset growth rates:

$$\Delta \log RK_{it}^{NA} = \sum_a \bar{v}_{ait} \Delta \log K_{ait}, \tag{C8}$$

with $\bar{v}_{ait} = \frac{1}{2}\left(v_{ait-1} + v_{ait}\right)$ and $v_{ait} = P_{ait}K_{ait} \Big/ \sum_a P_{ait}K_{ait}$. So the growth of the capital stock at constant national prices for each assets is weighted by its two-period average share in the capital stock at current national prices. Equation (C8) defines the growth rate of $RK^{NA}$, with 2005=1.

### v) Capital stock at current reference prices

The computation of the capital stock at current (reference) prices involves converting the capital stock at current national prices using a price level for the capital stock. Specifically, in terms of PWT variables, the current-price capital stock is defined as:

$$CK_{it} = \sum_a K_{ait} \Big/ \mathcal{E}_{it} PL\_K_{it}, \tag{C9}$$

where $K_{ait}$ is the capital stock of asset $a$ at national prices and $\mathcal{E}_{it}$ is the nominal exchange rate relative to the US $. The relative price of the capital stock $PL\_K_{it}$ is computed from investment prices for each of the assets – denoted by the vector $P_{it}^I$ – while using the capital stocks at current national prices as weights and using the U.S. as the comparison country 0:

$$PL\_K_{it} = \prod_{j=1}^{C} \left(P_{ij,t}^F P_{j0,t}^F\right)^{1/C} \Big/ \mathcal{E}_{it}, \text{ where } P_{ij,t}^F = \left[\left(\frac{P_{it}'^I K_{it}}{P_{jt}'^I K_{it}}\right)\left(\frac{P_{it}'^I K_{jt}}{P_{jt}'^I K_{jt}}\right)\right]^{1/2}. \tag{C10}$$

The prices for each asset are based on PPP benchmark surveys: the six ICP surveys since 1970 and the more frequent surveys by the OECD and Eurostat since 1995. The prices from these surveys do not directly map into the six assets we use. In some surveys, there would not be sufficiently detailed data to separately distinguish each of the six assets. At a minimum, it is always possible to distinguish investment in structures from investment in machinery and equipment and usually this latter category is also split between transport equipment and other machinery. In case of missing detail, the more aggregate price is applied to each of the detailed assets, such as the machinery price for ICT assets and for other machinery. There would also

sometimes be more detailed PPPs than the six we need and in those cases, we use a GEKS procedure with investments from the surveys as weights to arrive at the required six asset prices. We follow the same procedure as discussed in Appendix C to estimate a full panel of investment prices: using benchmark prices where available, interpolating between benchmarks and extrapolating if no earlier or later benchmark is available.

Figure C4 provides a cross-country view for 2005 of the resulting relative price level of the capital stock. As Hsieh and Klenow (2007) found, the prices of investment goods in poorer countries are high relative to the price of consumption and this is confirmed for 2005 in the left panel of Figure C4. Since many investment goods are traded, their prices are relative close to the exchange rate. In contrast, a considerable part of consumption is non-traded and prices in the non-traded sector tend to be lower in poorer countries. However, structures are non-traded, so their prices will be more similar to consumption prices. Since the depreciation rate of structures is lower than of other assets, the weight of structures in the capital stock will be larger than in total investment. As a result, the price level of the capital stock gives greater weight to the non-traded part of investment than the price level of investment. As a result, the right panel of Figure C4 shows a relationship between the relative price of the capital stock and $CGDP^o$ per capita that is significantly smaller than between the relative investment price and $CGDP^o$ per capita. It follows that the capital stock levels in PWT8 will typically be higher in poorer countries than the single-asset capital stock estimates that have been predominantly used in the literature.

The reason is that those single-asset capital stocks were constructed using PWT investment series, without taking the larger weight of structures prices in the relative capital stock into account. We constructed a single-asset capital stock using the same procedure as in Caselli (2005), so cumulating PPP-converted investment assuming a common depreciation rate

**Figure C4: The Investment and Capital Stock Price Relative to the
Household Consumption Price Level in 2005**
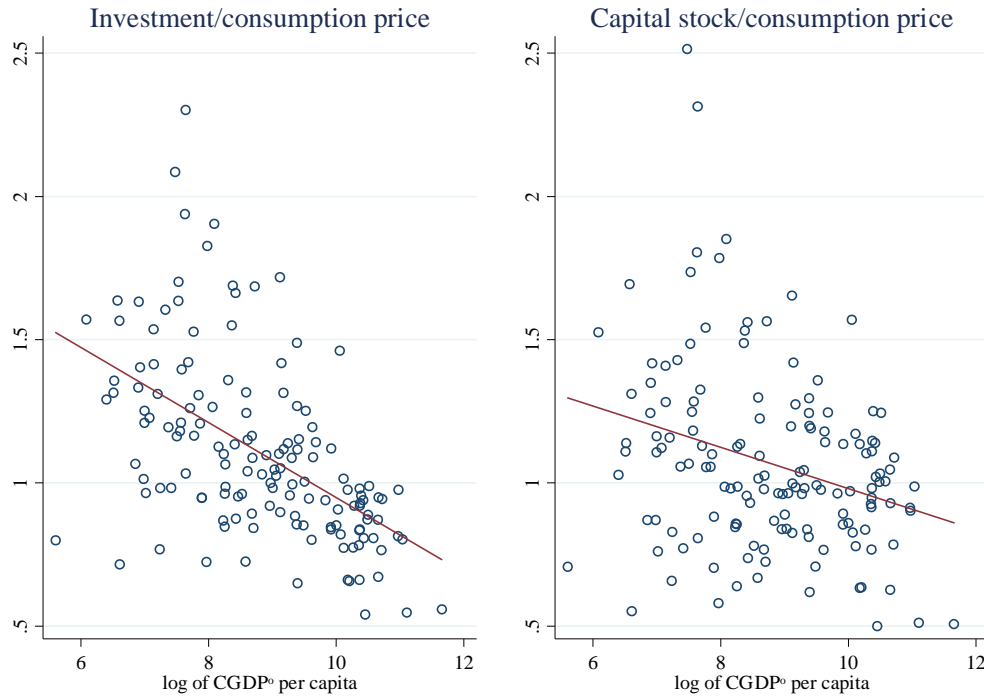


Note: only countries that participated in the 2005 ICP benchmark survey are included. The price level of investment goods (left panel) and the price level of the capital stock (right panel) is divided by the price level of household consumption.

(4%, based on Figure C3). The differences in 2005 between this single-asset capital stock and the PWT8.0 capital stock are shown in Figure C5. As the figure shows, the differences are quite large and, as predicted from Figure C4, the single-asset capital stock significantly underestimates capital input in poorer countries. The PWT8 capital stocks thus provide a different view than earlier approaches and a view that is more closely linked to the concept it represents.

### *Labor shares*

This section is devoted to estimating the share of labor income in GDP, the *LABSH* variable in PWT8. This is a challenge because, in contrast to the labor income of employees, the labor income of self-employed workers is not directly observable as their income consists of

**Figure C5: The Difference Between the Single-Asset and PWT8 Capital/Output ratio in 2005**
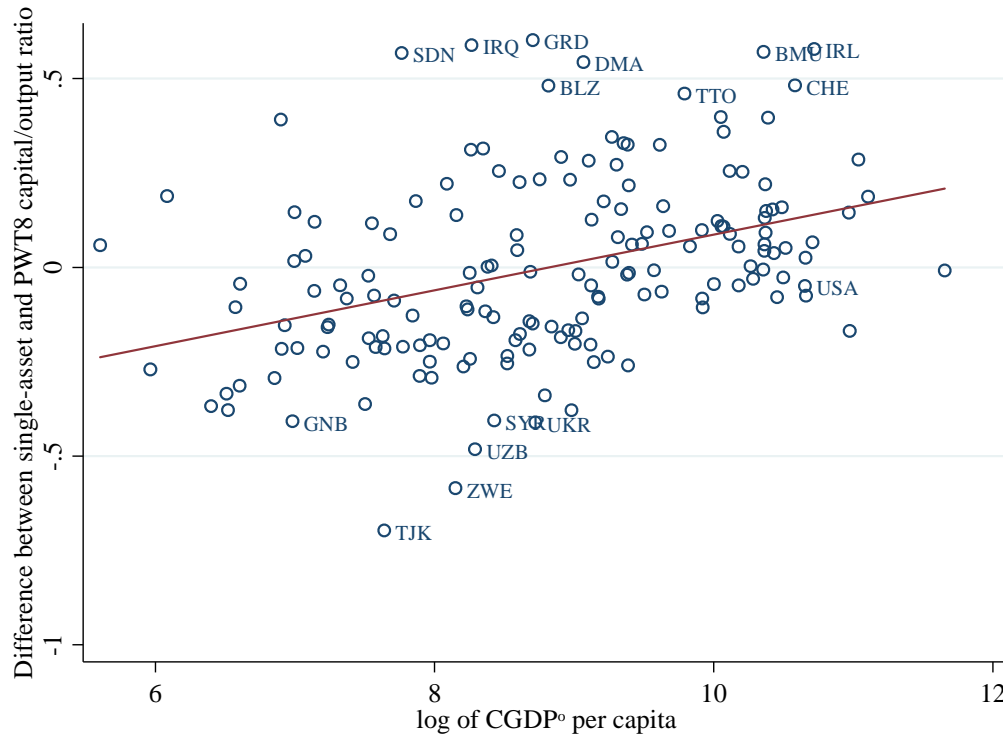


Note: single-asset capital stock cumulates overall real investment and depreciates it at a common 4 percent. The PWT8 capital stock uses asset-specific investment and depreciation rates and converts to a common currency using a capital stock PPP.

compensation for both their labor supply and any capital they may own. This issue was taken up by Gollin (2002), who discusses different methods for estimating the labor compensation of self-employed workers. He showed for a modest set of countries that suitably adjusted labor shares are much more similar than naïve shares that ignore the labor compensation of the self-employed. This is because in poorer countries, more people are self-employed, which compensates for the lower naïve labor share in poorer countries.

For PWT8 we build upon these efforts by adding an adjustment method, increasing the range of countries, and, most importantly, extending the time period covered. Based on these, we come up with a 'best estimate' labor share based that is subsequently used in our TFP

calculations. The end result is labor share estimates for up to 127 out of 167 countries in PWT8, covering the period since 1950.

### i) *Basic data and adjustment methods*

The starting point is National Accounts data on compensation of employees, GDP at basic prices[12] and mixed income. Mixed income is the total income earned by self-employed workers, so it is a combination of capital and labor income. Given the aim of dividing the income of the self-employed between labor and capital, data on mixed income is very helpful by providing an upper bound to the amount of labor income earned by the self-employed. Two of Gollin's (2002) adjusted labor shares rely on mixed income information: the first allocates all mixed income to labor, assuming that self-employed workers only use labor input; while the second assumes that self-employed workers use labor and capital in the same proportion as the rest of the economy.

Mixed income data is available for 60 of the countries in PWT, so additional information is required. Gollin's (2002) third approach additionally uses data on the number of employees and the number of self-employed and assumes that self-employed earn the same average wage as employees. These data are drawn from the ILO LABORSTA database.[13] The 'same-wage' assumption may not be too far off the mark in advanced economies where the share of employees in the total number of persons engaged (employees + self-employed) is 85-95 percent. However, in many emerging economies this share is below 50 percent and as low as 4 percent.[14] In those countries, using information on the wages of employees will overstate the labor income of the self-employed.

---

[12] Net taxes on products are excluded since this is not income accruing to any of the factor inputs but a direct transfer to the government.

[13] These data are supplemented by data from the Socio-Economic Accounts of the World Input-Output Database, see www.wiod.org.

[14] The share of employees in persons engaged is strongly positively related to GDP per capita.

We therefore propose an alternative adjustment method. Most self-employed workers are active in agriculture. According to the Socio-Economic Accounts (SEA) of the World Input-Output Database (see Timmer, 2012), agriculture employs about half of the self-employed in poorer countries. The agricultural sector also uses very few fixed assets in these countries as, according to the SEA, the agricultural labor share (accounting for the labor income of self-employed) is over 90 percent of value added, on average. We therefore have an adjusted labor share that adds all of value added in agriculture to labor compensation of employees. This adjustment could be too large as it ignores all income from capital and land and the labor income of employees in this sector is double-counted. On the other hand, the labor income of self-employed outside agriculture is ignored.

This provides us with data on the share of labor compensation of employees in GDP at basic prices and four adjusted shares, namely: two based on mixed income, one based on the share of employees in the number of persons engaged and one based on the share of agriculture in GDP. Table C5 shows descriptive statistics for these shares in 2005. Statistics for a single year makes it easier to illustrate the cross-country variation since the basic data for some countries is much more extensive than for others and 2005 gives the largest country coverage.[15] Table C5 shows that, as in Gollin (2002), the naïve approach of using labor compensation of employees leads to very low labor shares of 42 percent on average and as low as 5 percent (in Nigeria). There are also some very high labor shares, however, up to 89 percent (in Bhutan). Adjustments 1 and 2 – based on mixed income data– show notably higher labor shares, though these can only be computed for 53 of the 108 countries. Especially some of the main oil- producing countries (Qatar, Oman, Venezuela) show quite low labor shares (20-45%) based on these adjustments.

---

[15] The sample of countries is different in each row. Descriptive statistics for the common sample of 42 countries shows very similar averages but with a smaller range.

**Table C5: Descriptive Statistics for the Different Labor Share Alternatives, 2005**

| Share | Mean | StDev | Min | Max | Obs |
|---|---|---|---|---|---|
| Naïve share | 0.42 | 0.14 | 0.05 | 0.89 | 108 |
| Adjustment 1, mixed income | 0.60 | 0.12 | 0.20 | 0.90 | 53 |
| Adjustment 2, part mixed income | 0.53 | 0.13 | 0.18 | 0.73 | 53 |
| Adjustment 3, average wage | 0.66 | 0.24 | 0.22 | 2.27 | 74 |
| Adjustment 4, agriculture | 0.51 | 0.14 | 0.17 | 1.13 | 108 |

Note: StDev: standard deviation, Obs: number of observations. Naïve share is the share of labor compensation of employees (COMP) in GDP. Adjustment 1 adds mixed income (MIX): (COMP+MIX)/GDP. Adjustment 2 assumes the same labor share for mixed income as for the rest of the economy: COMP/(GDP-MIX). Adjustment 3 assumes the same average wage for self-employed (SEMP) as for employees (EMPE): (COMP/GDP)*(EMPE+SEMP)/EMPE. Adjustment 4 adds value added in agriculture (AGRI): (COMP+AGRI)/GDP.

Using information on the number of self-employed and assuming they earn the same average wage as employees (adjustment 3) leads to average labor shares that are close to the commonly assumed labor share of 0.7 in Caselli (2005) and many others. Here too, though, there are countries with very low shares (Kuwait: 0.22), and some with unrealistically high shares, such as Bhutan (2.27). Bhutan already had a very high labor share according to the naïve share, so this overestimation is to be expected. This could indicate that, in contrast to National Accounting rules, the statisticians in Bhutan have already imputed the labor income of self-employed in their 'employee compensation' numbers. Lesotho's labor share under adjustment 3 of 2.05 indicates a similar problem.

The fourth adjustment adds the value added share of agriculture to the naïve share. The average share is somewhat lower than the commonly assumed two-thirds labor share but there are no countries with labor shares as extreme as under adjustment 3, though Bhutan is again the country with the highest labor share. In this broad cross-country setting, it would seem that any of the four adjustments would count as an improvement over the naïve share, but also that the mechanical application of one of these adjustments would not fit all countries equally well. This is even more apparent when comparing the cross-country pattern of adjustments 3 and 4, as done

in Figure C6. This shows that adjustment 3 leads to some very high labor shares for the poorer

countries. Indonesia (IDN), for instance, shows a labor share of 92 percent under adjustment 3

but a share of only 44 percent under adjustment 4. The 92 percent adjustment is almost certainly

too high since more than 80 percent of Indonesia's self-employed work in agriculture or

distributive trade (based on SEA data). Adding the full value added earned in those sectors as

labor compensation would lead to a labor share of 60 percent. Indonesia's remaining self-

employed are unlikely to earn another 30 percent of GDP.

**Figure C6: Labor Shares vs. GDP per capita in 2005, Adjustments 3 and 4**
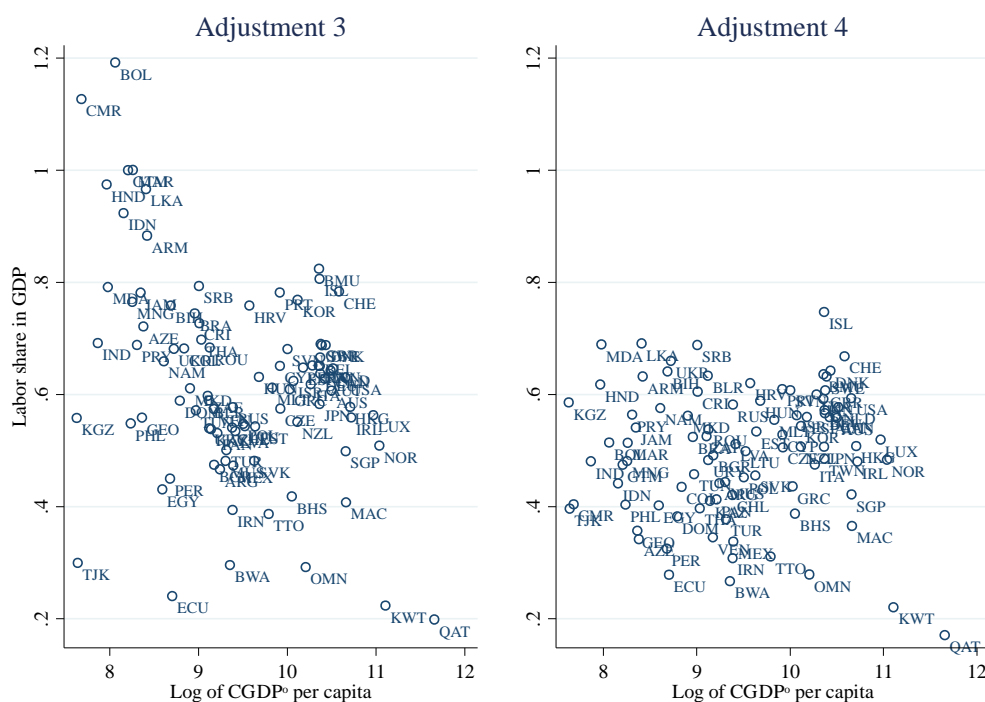


Note: see notes to Table C5 for a description of the adjustments. Both figures include only countries for which data on adjustment 3 is available and excludes Bhutan and Lesotho (both with labor shares under adjustment 3 exceeding 2).

## ii) Best estimate labor share

These results suggest that a single adjustment approach is not appropriate for all countries. We therefore construct a 'best estimate' labor share based on the following three rules:

1. Where available, adjustments based on mixed income seem preferable as this income directly relates to the income of self-employed, giving an upper-bound to the labor share. Adjustment 1 seems fairly extreme by assuming that self-employed use no capital at all, so we consider adjustment 2 to be the more plausible approach and use those labor shares if available.

2. Whenever mixed income data is available in National Accounts statistics, the naïve labor share never exceeds 0.66. So if the naïve labor share is larger than 0.7 in a particular country, it seems reasonable, like in Bhutan, that this share already includes an imputation for self-employed labor income. In those cases, the naïve share is used directly.

3. Given the patterns shown in Figure C6, there seems to be a greater chance of overestimating the labor share than underestimating the labor share. A conservative estimate would thus be the smaller of adjustments 3 and 4. So this is what we use if there is no mixed income data and the naïve share is below 0.7.

To ensure complete coverage over the years, we assume labor shares remain constant or we linearly interpolate if there are missing years in the middle of the sample. After these interpolations and extrapolations, we apply the three rules. Table C6 summarizes the resulting 'best estimate' labor shares for 2005.

**Table C6: Summary Statistics of the 'Best Estimate' Labor Share in 2005,**
**By type of adjustment**

| Share | # of countries | Mean | StDev | Min | Max |
|---|---|---|---|---|---|
| Overall | 127 | 0.52 | 0.14 | 0.18 | 0.89 |
| *of which:* | | | | | |
| Adjustment 2, part mixed income | 60 | 0.52 | 0.13 | 0.18 | 0.73 |
| Adjustment 3, average wage | 4 | 0.40 | 0.16 | 0.24 | 0.58 |
| Adjustment 4, agriculture | 62 | 0.52 | 0.14 | 0.22 | 0.85 |
| Naïve share | 1 | 0.89 | | | |

Note: see notes to Table C5 for details on construction of the naïve and adjusted shares.

By interpolating and assuming shares constant over time, country coverage increases to 127 countries (out of 167 in PWT8). The resulting cross-country average of 0.52 is lower than Gollin's (2002) preferred 0.7 estimate, but it shows only a somewhat larger range than his of 0.34-0.91.[16] The average is lower, which is partly related to revisions of the underlying data for mixed income. Of the 15 countries with mixed income in both Gollin's data and here, the average 'adjustment 2' labor share is 68 percent in Gollin's (2002) data and 60 percent based on the current vintage of National Accounts data.[17] In contrast, the naïve share and 'adjustment 3' share are very similar for the overlapping set of countries. In addition, there seems to be a downward trend in labor shares over time, see below for more discussion.

Table C6 also illustrates that for almost half the countries, information on mixed income is available and therefore used. Adjustment 4 is used for most other labor share estimates and Adjustment 3 and the naïve share are only used for a few countries. The overall pattern is very similar if labor shares that are interpolated or assumed constant are dropped from the sample. The patterns shown in Table C6 for labor shares in 2005 also hold for the full sample, though there are many fewer labor shares based on observed data before the 1990s.

---

[16] This is the range of his Table 2, not the 0.65-0.80 he mentions in his abstract. Only about half of his labor shares fall within the narrow 0.65-0.80 range.

[17] This compares shares for the same year as in Gollin (2002).

**Table C7: Labor Shares and Variation across Income Levels and Time**

| | (1) Naïve share | (2) Best estimate | (3) Best estimate | (4) Best estimate | (5) Best estimate |
|---|---|---|---|---|---|
| Log of CGDP$^o$ per capita | 0.0741*** (0.00870) | 0.00553 (0.00942) | 0.00591 (0.00798) | 0.0105 (0.00748) | -0.0276** (0.0113) |
| Oil country | | | -0.153*** (0.0290) | -0.146*** (0.0265) | |
| Time trend | | | | -0.00336*** (0.000495) | -0.00159*** (0.000461) |
| Country dummies | no | no | no | no | yes |
| Observations | 2975 | 2237 | 2237 | 2237 | 2237 |
| R-squared | 0.335 | 0.003 | 0.183 | 0.273 | 0.204 |

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors, clustered by country, in parentheses. Dependent variables are the naïve share (column (1)) or the best estimate labor share. Sample excludes labor shares that are interpolated or assumed constant from other years. Oil countries are OPEC countries and those countries for which the share of energy exports exceeds one-third. This share was chosen as all OPEC countries have an energy export share that is at least this large.

Table C7 analyses the cross-country patterns in the labor share data by relating labor shares to (the natural log of) CGDP$^o$ per capita levels, excluding any labor shares that are assumed constant or interpolated. The naïve share in column (1) shows a strong positive relationship with GDP per capita, as Gollin (2002) also observed in his smaller sample. When using the best estimate labor share, no significant relationship with income levels is found in column (2). What does matter substantially is whether it is an oil country – the OPEC members and any other country in which energy exports accounts for at least one-third of total exports. Column (3) shows that those countries have labor shares that are on average 15 percentage points lower. Column (4) adds a linear time trend, and this shows a significant decline in labor shares over time. This remains the case when including country dummies in column (5). In that specification, increases in income levels are even associated with *lower* labor shares, compared with the higher labor shares from column (1). The time trend is less steep in column (5), but still highly significant. Moreover, the pattern of declining labor shares is found across the whole sample of countries as there is a decline in the labor share in 89 of the 127 countries and the

trend is there for rich and poor countries alike.[18] Finally, if year dummies rather than a linear

time trend is used, the assumed linear relationship over time of Table C7 is confirmed.

These results pose the important question how to explain this variation. Karabarbounis

and Neiman (2014) propose cheaper investment goods to explain declines in labor shares over

time, but other channels have been highlighted as well: capital-biased technological change,

international trade (Harrison, 2005) and increased financialisation. This is fruitful area for

additional research. Regardless of the underlying cause, though, this analysis illustrates quite

clearly that the standard 'one-size-fits-all' labor share of 70 percent that is commonly used in the

literature is a simplification that is not supported by the facts.

**Appendix D: Extrapolating PPPs and Biased Balassa-Samuelson estimates**

To see why extrapolating PPPs using relative inflation could lead to systematically biased

PPP estimates, we use the following stylized example from Deaton (2012). We use the

expenditure function to measure real GDP on the expenditure side and consider a two-country

setting where consumers in both countries have common but non-homothetic tastes. The rate of

inflation in each country can be measured by totally differentiating the expenditure function

(obtaining a continuous-time Divisia index):

$$d \ln P_{jt} = \left( \frac{\partial \ln E\left( p_{jt}, u_j \right)}{\partial p_{jt}} \right)' d \ln p_{jt} = \ln s'_{jt} d \ln p_{jt}, \quad j = 1, 2, \qquad \text{(D1)}$$

where $p_{jt}$ is the vector of prices in country $j$, $s_{jt}$ is the vector of the budget shares in country $j$, $E$ is

the expenditure function, and $P_{jt}$ is the overall price level. The shares in (D1) must be evaluated

at current prices, so that in-between two prices $p_{jt-1}$ and $p_{jt}$ these shares will be changing.

Suppose that we approximate the discrete change in (D1) between these two price vectors by the

---

[18] Though the trend is less steep and less significant in richer countries.

Törnqvist index,

$$\Delta \ln P_{jt} \approx \tfrac{1}{2}(\ln s_{jt-1} + \ln s_{jt})\Delta \ln p_{jt}, \quad j = 1,2.$$

Then the inflation differential between the two countries can then be written as:

$$\Delta \ln P_{2t} - \Delta \ln P_{1t} \approx \tfrac{1}{2}(\ln s_{2t-1} + \ln s_{2t})'\Delta \ln p_{2t} - \tfrac{1}{2}(\ln s_{1t-1} + \ln s_{1t})'\Delta \ln p_{1t} \quad (D2)$$

Next, suppose that we approximate the PPP index of country 2 relative to country 1 by a Törnqvist price index, computed using the shares at time $t$:

$$\ln PPP_{21,t} \approx \tfrac{1}{2}\big(s_{2t} + s_{1t}\big)'\big(\ln p_{2t} - \ln p_{1t}\big). \quad (D3)$$

This PPP index relies on budget shares for the two countries, just like the Fisher indexes from section 3 were based on the quantity vectors of both countries; in both cases this allows for substitution in response to price differences. (D3) is an approximation because, with non-homothetic tastes, the Törnqvist price index does not precisely measure the ratio of expenditures in the two countries needed to obtain the same level of utility. Combining (D2) and (D3), we get the following expression for the change in the PPP index over time:

$$\Delta \ln PPP_{21,t} \approx \tfrac{1}{2}\big(s_{2t} + s_{1t}\big)'\big(\ln p_{2t} - \ln p_{1t}\big) - \tfrac{1}{2}\big(s_{2t-1} + s_{1t-1}\big)'\big(\ln p_{2t-1} - \ln p_{1t-1}\big)$$

$$= \tfrac{1}{2}(s_{2t-1} + s_{2t})'\big(\ln p_{2t} - \ln p_{2t-1}\big) - \tfrac{1}{2}(s_{1t-1} + s_{1t})'\big(\ln p_{1t} - \ln p_{1t-1}\big)$$
$$- \tfrac{1}{2}\big(s_{2t} - s_{1t}\big)'\big(\Delta \ln p_{2t} + \Delta \ln p_{1t}\big) + \tfrac{1}{2}\Delta s_{2t}'\Delta \ln p_{2t} - \tfrac{1}{2}\Delta s_{1t}'\Delta \ln p_{1t} \quad (D4)$$

$$\approx \underbrace{\big(\Delta \ln P_{2t} - \Delta \ln P_{1t}\big)}_{\text{Inflation differential}} - \underbrace{\tfrac{1}{2}\big(s_{2t} - s_{1t}\big)'\big(\Delta \ln p_{2t} + \Delta \ln p_{1t}\big)}_{\text{Bias}}$$

The first approximation in (D4) is the same as that in (D3); the following equality is obtained with detailed algebra; and the final approximation follows from (D2) and by treating the final term $(\tfrac{1}{2}\Delta s_{2t}'\Delta \ln p_{2t} - \tfrac{1}{2}\Delta s_{1t}'\Delta \ln p_{1t})$ as small. Then (D4) states that the inflation differential

between the two countries is a potentially biased estimator for the change in the PPP index, with the bias indicated by the final term on the right-hand side.

We can infer the direction of the bias when comparing rich and poor countries by considering that inflation of traded products (with higher productivity growth) tends to be lower than inflation of non-traded products. Suppose country 2 is poorer than country 1. In poorer countries traded products – such as food – generally have larger budget shares. As a result, there is a negative correlation between the components of the vectors of budget share differences $(s_{2t} - s_{1t})$ in (D4) and average inflation $\frac{1}{2}(\Delta \ln p_{2t} + \Delta \ln p_{1t})$. Therefore, the final term on the right-hand side of equation (D4) is negative, and including the minus sign will make it positive, so that the overall bias is positive. It follows that the PPP of country 2 relative to country 1 will increase at a *faster rate* than implied by the difference in overall inflation.

This result has important implications for the Balassa-Samuelson effect based on PPPs extrapolated using relative inflation rates. Since PPPs are typically extrapolated backwards in time – as recent ICP surveys cover more countries than the earlier ones – using the difference in relative inflation rates would overstate the 'true' PPP for poor countries if the previous argument holds. This overstatement would raise the price level of the poorer countries, and therefore bias the Balassa-Samuelson regression coefficient towards zero. The further back in time, the larger the bias would be, which explains why the Balassa-Samuelson effect is not found in our estimates when using extrapolated price levels.

To more firmly establish this finding that was already suggested in Figure 3 in the main text, consider Figure D1, which plots the annual coefficient $\beta_{1t}$ from the Balassa-Samuelson regression:

$$\ln\left(\frac{PPP_{it}^o}{\mathcal{E}_{it}}\right) = \beta_0 + \beta_{1t}\ln\left(\frac{CGDP_{it}^o}{POP_{it}}\right) + \varepsilon_{it}. \qquad (D5)$$

The figure shows a qualitatively similar pattern as Bergin *et al* (2006), with a low and insignificant coefficient in the early years, which rises steadily and becomes highly significant in later years. The main substantive difference is that it takes until 1974 before the slope coefficient in (23) turns significantly positive, versus the mid-1950s in Bergin *et al* (2006).[19] This is a puzzling finding, as the Balassa-Samuelson effect was already identified in data for the 1950s

**Figure D1: The Balassa-Samuelson Effect in PWT8.1, 1950-2011**
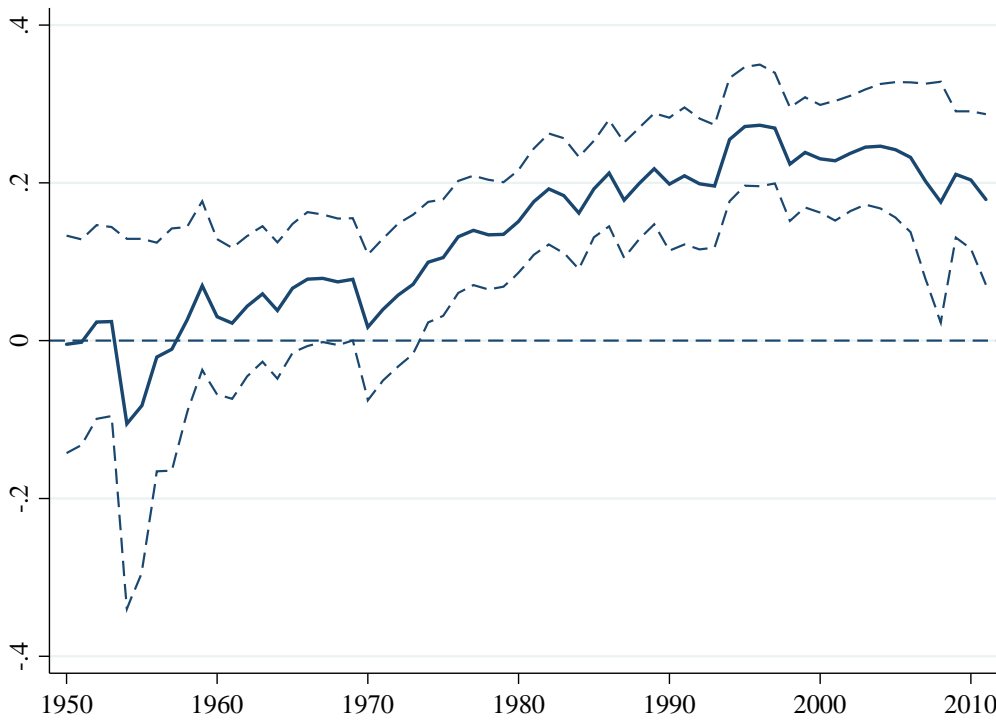


**Notes:** The figure plots $\beta_{1t}$ from equation (D5) and its 95% confidence interval for all observations in PWT8.1, excluding those for which, due to extreme swings, no market exchange rate is available. Each $\beta_{1t}$ is from a cross-sectional regression for all observations in year *t*.

---

[19] The same analysis based on PWT7.1 shows a qualitatively similar pattern, but the coefficient does not turn significantly positive until 1988.

and 1960s in e.g. Balassa (1964), therefore raising the question why the effect shows up so much later in PWT8.

To determine whether the non-benchmark observations are in fact leading to the disappearance of the Balassa-Samuelson effect in early years, we distinguish between PPP benchmarks and observations interpolated between those benchmarks on the one hand, and observations that are extrapolated on the other hand. Figure D2 shows the result from the same cross-sectional regressions as in Figure D1, but run separately on the two subsamples from 1970-2005; note that there are no benchmark observations before 1970 and that the last global benchmark comparison was in 2005. Figure D2 shows that in the *benchmark/interpolated* sample there is a consistently positive and significant coefficient $\beta_{1t}$ of 0.3 on average, confirming the Balassa-Samuelson effect, while in the *extrapolated* sample, the coefficient is rarely significantly different from zero. This confirms the hypothesis that the extrapolation procedure biases the Balassa-Samuelson effect towards zero.[20] The pattern in Figure D2 can thus be best understood as a weighted average between a positive, significant coefficient in the benchmark/interpolated sample and a zero coefficient in the extrapolated sample. As the weight of the benchmark/ interpolated sample increases over time, the overall coefficient increases and turns significant. The study of Bergin *et al* (2006) is based on PWT 6 which relied exclusively on extrapolation of PPPs from a recent benchmark year which would explain their findings. Our findings illustrates the usefulness of including historical benchmark material and clearly distinguishing between benchmark/interpolated and extrapolated observations, as done in PWT8.

---

[20] If we restrict ourselves to the countries and years included in the benchmark-or-interpolated sample and compute extrapolated PPPs, we still find that the Balassa-Samuelson coefficient $\beta_{1t}$ is biased towards zero.

**Figure D2: The Balassa-Samuelson Effect in PWT8.0 for Benchmark or Interpolated and for Extrapolated Observations, 1970-2005**
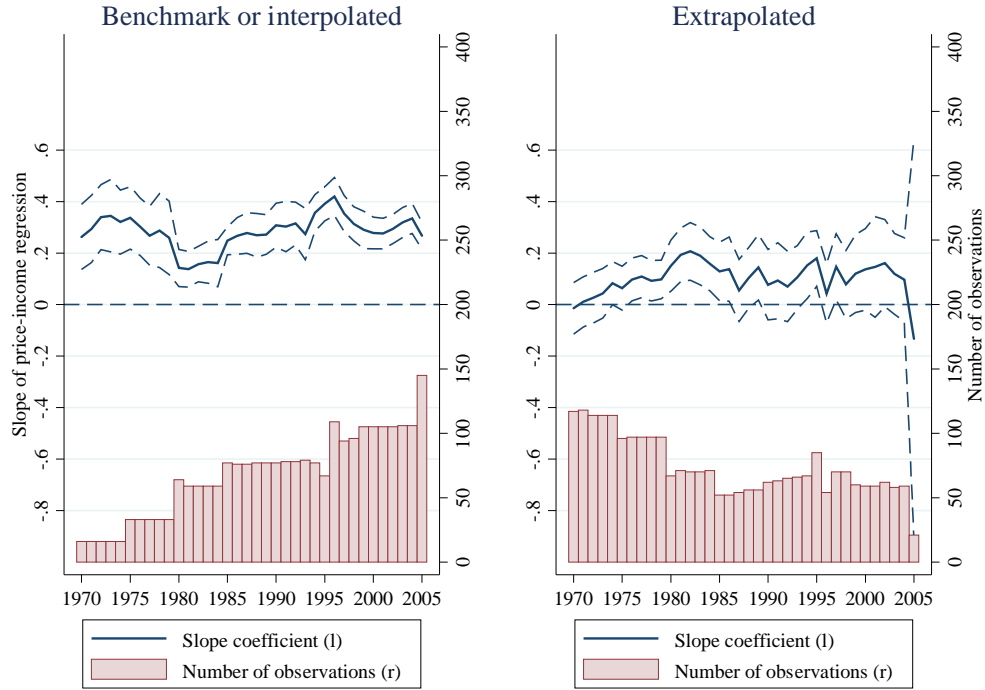


**Note**: The figure plots $\beta_{1t}$ from (D5) and its 95% confidence interval for all observations in PWT8.0. Each $\beta_{1t}$ is from a cross-country regression in year $t$. The left-hand panel includes only observations from PPP benchmarks or interpolated between PPP benchmarks. The right-hand panel includes only observations extrapolated from PPP benchmarks using inflation rates, i.e. all other observations. Excluded are those observations for which, due to extreme swings, the market exchange rate is replaced in PWT8.1 by an estimated exchange rate based on a relative PPP assumption.

**References:**

Balassa, Bela. 1964. "The Purchasing-Power Parity Doctrine: A Reappraisal." *Journal of Political Economy* 72 (6): 584–96.

Barro, Robert and Jong-Wha Lee. 2013. "A New Data Set of Educational Attainment in the World, 1950-2010." *Journal of Development Economics* 104: 184–98.

Bergin, Paul R., Reuven Glick and Alan M. Taylor. 2006. "Productivity, Tradability, and the Long-Run Price Puzzle." *Journal of Monetary Economics* 53 (8): 2041–66.

Caselli, Francesco. 2005. "Accounting for Cross-Country Income Differences." In Phillipe Aghion and Steven N. Durlauf, editors, *Handbook of Economic Growth, Volume 1A,* 679-741. Amsterdam: Elsevier,.

Caselli, Francesco and Daniel J. Wilson. 2004. "Importing Technology." *Journal of Monetary Economics* 51 (1): 1–32.

Deaton, Angus. 2012. "Consumer Price Indexes, Purchasing Power Parity Exchange Rates, and Updating." Unpublished paper for the PWT Workshop May 2012, Philadelphia.

Diewert, W. Erwin. 2008. "The Economic Theory of Index Numbers: A Survey." In Angus Deaton, editor, *Essays in the Theory and Measurement of Consumer Behaviour in Honour of Sir Richard Stone*, 163–208. Cambridge: Cambridge University Press.

Feenstra, Robert C., Alan Heston, Marcel P. Timmer, and Haiyan Deng. 2009. "Estimating Real Production and Expenditures Across Countries: A Proposal for Improving the Penn World Tables." *Review of Economics and Statistics* 91 (1): 201–12.

Feenstra, Robert C. and John Romalis. 2014. "International Prices and Endogenous Quality." *Quarterly Journal of Economics* 129 (2): 477–527.

Fraumeni, Barbara M. 1997. "The Measurement of Depreciation in the U.S. National Income and Product Accounts." Survey of Current Business. July: 7-23.

Gollin, Douglas. 2002. "Getting Income Shares Right." *Journal of Political Economy* 110 (2): 458–74.

Harberger, Arnold. 1978. "Perspectives on Capital and Technology in Less Developed Countries." In Michael J. Artis and Avelino R. Nobay, editors, *Contemporary Economic Analysis*, London: Croom Helm.

Harrison, Ann. 2005. "Has Globalization Eroded Labor's Share? Some Cross-Country Evidence." MPRA Paper No. 39649.

Hill, Robert J. 2004. "Constructing Price Indexes across Space and Time: The Case of the European Union." *American Economic Review* 94 (5): 1379–410.

Hsieh, Chang-Tai and Peter J. Klenow. 2007. "Relative Prices and Relative Prosperity." *American Economic Review* 97 (3): 562–85.

Inklaar, Robert. 2010. "The Sensitivity of Capital Services Measurement: Measure All Assets and the Cost of Capital." *Review of Income and Wealth* 56 (2): 389–412.

Inklaar, Robert and Marcel P. Timmer. 2013. "Capital and TFP in PWT8.0." mimeo, available at: www.ggdc.net/pwt.

Karabarbounis, Loukas and Brent Neiman. 2014. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129 (1): 61–103.

Malmquist, Sten. 1953. "Index Numbers and Indifference Surfaces." *Trabajos de Estatistica* 4: 209–42.

Nehru, Vikram, and Ashok Dhareshwar. 1993. "A New Database on Physical Capital Stock: Sources, Methodology and Results." *Revista de Analisis Economica* 8 (1): 37–59.

Rao, D. S. Prasada, Alicia Rambaldi and Howard Doran. 2010. "Extrapolation of Purchasing Power Parities using Multiple Benchmarks and Auxiliary information: A New Approach." *Review of Income and Wealth* 56 (S1): S59–98.

Timmer, Marcel. 2012. "The World Input-Output Database (WIOD): Contents, Sources and Methods." WIOD Working Paper Number 10, downloadable at (May 22, 2014): http://www.wiod.org/publications/papers/wiod10.pdf .