Cross-country income levels over time

Inklaar, Robert; Prasada Rao, D. S.

Published in:
American Economic Journal. Macroeconomics

DOI:
10.1257/mac.20150155

American Economic Journal: Macroeconomics 2017, 9(1): 265–290
https://doi.org/10.1257/mac.20150155

Cross-Country Income Levels over Time: Did the Developing World Suddenly Become Much Richer?[†]

By Robert Inklaar and D.S. Prasada Rao*

The latest global survey on relative prices and income levels, for the year 2011, showed changes to relative income levels that were larger in lower income countries, thereby narrowing the world income distribution compared to estimates based on the previous, 2005, survey. This paper examines whether changes in measurement methodology between the 2005 and 2011 survey can explain these large differences. We construct a counterfactual set of relative prices for 2005 that harmonizes measurement, and we no longer find systematic differences across income levels, implying that international income inequality based on the 2005 survey was overstated. (*JEL* C82, D31, E01, E23, E31, O11)

Our understanding of differences in living standards across countries, international income inequality, and global poverty relies heavily on the periodical global price comparisons of the International Comparison Program (ICP).[1] Without ICP relative prices, comparisons of income levels across countries would have to rely on exchange rates, which systematically underestimate the purchasing power of consumers in low-income countries.[2] But despite the conceptual appeal of ICP relative prices, they are often also a source of controversy. For instance, Deaton (2010) showed how results from consecutive ICP rounds, up to and including ICP 2005, led to upward revisions in income inequality across countries. Similarly, Ciccone and Jarociński (2010) and Johnson et al. (2013) show how various results from the

*Inklaar: University of Groningen, Faculty of Economics and Business, P.O. Box 800, 9700 AV Groningen, Netherlands (e-mail: R.C.Inklaar@rug.nl); Rao: The University of Queensland, School of Economics, Colin Clark Building, St Lucia, Brisbane Qld 4072, Australia (e-mail: d.rao@uq.edu.au). A previous version of this paper was circulated under the title "Cross-country income levels of time: can ICP 2005 and ICP 2011 be reconciled?" The authors would like to thank four anonymous referees, Angus Deaton, Sergey Sergeev, Arvind Subramanian, Marcel Timmer, and participants at the World Bank poverty measurement workshop of January 2015, the ICP workshop in Prague of May 2015, and seminars at the University of Graz and the University of Groningen for helpful comments; and the World Bank for providing detailed researcher data underlying ICP 2005 and ICP 2011. This work was undertaken when Rao was a visiting researcher at the Groningen Growth and Development Centre in August 2014 and at UNU-MERIT during October 2014. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

[†] Go to https://doi.org/10.1257/mac.20150155 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] See, for example, Anand and Segal (2008), Deaton (2010), Deaton and Heston (2010), Chen and Ravallion (2010b), and Milanovic (2012). See also Feenstra, Inklaar, and Timmer (2015) on how the ICP results are a crucial input in the Penn World Table and World Bank (2013) for details on ICP.

[2] The Balassa (1964)-Samuelson (1964) effect, whereby prices of non-traded products tend to be lower in lower-income countries.

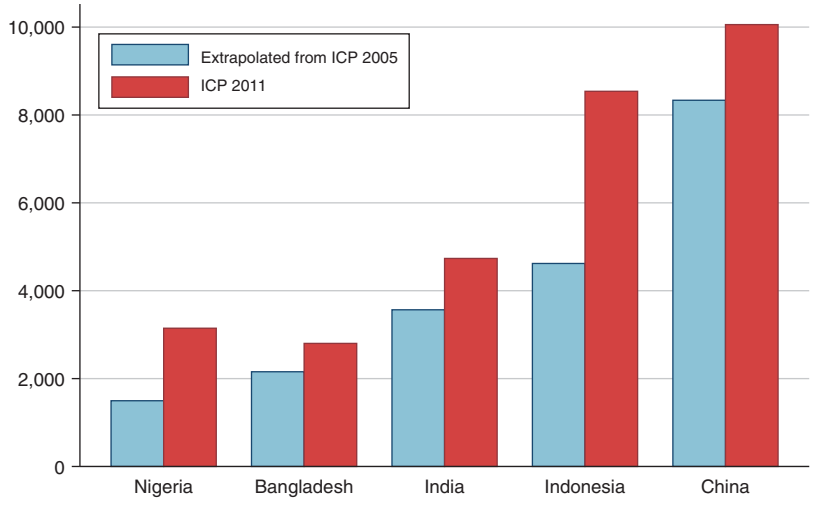

FIGURE 1. GDP PER CAPITA IN 2011 FOR SELECTED COUNTRIES BASED ON ALTERNATIVE RELATIVE PRICES (*in US dollars*)

Note: GDP per capita at current national prices and ICP 2011 relative prices from World Bank (2014); ICP 2005 relative prices (from World Bank 2008) extrapolated using the change in the country GDP deflator relative to the US GDP deflator.

Source: Computations based on World Bank (2008, 2014) and World Development Indicators

cross-country growth literature change depending on the vintage of relative price and income data that is used.[3] More broadly, changing the methods and models used to measure relative prices and income levels can have a notable effect on the resulting relative income estimates, as shown by Neary (2004); Almås (2012); and Feenstra, Inklaar, and Timmer (2015).

The most recent controversy followed the publication of findings from ICP 2011 (World Bank 2015). Compared with earlier estimates based on extrapolations from ICP 2005 (World Bank 2008), the average country saw its relative income and consumption levels increase—relative to the United States—by about 25 percent. Changes in developing countries were particularly large, resulting in major changes to the economic geography of the world and a narrowing of international income differences. Given the size and systematic pattern of these differences, they can have far-reaching consequences, ranging from a different number and geographical distribution of the world's poor to a different view on income convergence or appropriate policy.[4] It is thus no surprise that the revisions following ICP 2011 have been the topic of substantial commentary and debate.[5]

Figure 1 provides an illustration of these changes for a selection of countries based on GDP per capita for the year 2011. The first set of bars shows income levels

[3] Differences between data vintages in these studies reflect revisions to relative prices—the focus of our paper—and revisions to national GDP statistics—a source of differences we exclude.

[4] See Chen and Ravallion (2010b) on the impact of ICP 2005 on poverty; Pritchett and Summers (2014) on growth and convergence; and Aghion, Akcigit, and Howitt (2014) on productivity levels and appropriate policy.

[5] See Dykstra, Kenny, and Sandefur (2014); Kharas and Chandy (2014); Deaton and Aten (2014); and Ravallion (2014).

based on extrapolated relative prices from ICP 2005, while the second set of bars shows income levels based on ICP 2011. The differences can be very large, with Indonesia's GDP per capita increasing from $4,620 to $8,539. Even the rankings can change, as Nigeria used to have an income level that was 30 percent below that of Bangladesh, but, based on ICP 2011, it has an income level that is 12 percent higher. As we show later in this paper, differences of this magnitude are common.[6]

The periodic benchmark price comparisons from the ICP are critical for the compilation of the Penn World Table (PWT) and for the World Bank's World Development Indicators, which are an important source for researchers and analysts interested in income comparisons. If the differences between benchmarks are large and systematic, as is the case with the ICP 2011 and ICP 2005 benchmarks, this calls into question the common practice of using a single set of cross-country relative prices to "backcast" relative income levels across many years,[7] and the research results that depend on such data.[8]

Such revisionism is not yet called for because there were major changes in the measurement methods in ICP 2011 compared with ICP 2005.[9] A particular concern, as identified by Deaton and Aten (2014), is the linking of regions. In both ICP 2005 and 2011, prices were first collected and compared across the countries within a region and in the second stage, the regions were linked to allow for global price comparisons. Some regions consist mostly of low- and middle-income countries (Africa, Asia-Pacific) and some mostly of high-income countries (Eurostat/OECD), which means that changes to the linking approach can shift the prices of lower income countries relative to higher income countries. Since the linking approach was considerably improved and refined in 2011,[10] Deaton and Aten (2017) conclude that the linking data and methods are a good starting point for trying to understand why the results from ICP 2011 were so different from the ICP 2005 results.

The goal of this paper is to quantify the impact of the changes in measurement in ICP 2011 and examine if it is possible to reconcile and explain the changes to relative prices, and thus to income levels, from ICP 2005 and ICP 2011. To this end, we construct a counterfactual price comparison for 2005 using ICP 2011 methods and adjusting for linking biases in ICP 2005. We draw upon the diagnostic work of Deaton and Aten (2017), but provide a more comprehensive and quantitative assessment of the impact of the measurement changes.

To construct our counterfactual price comparison for 2005, we make use of the same detailed price and expenditure data that were used in ICP 2005 but implement the methods used in the compilation of ICP 2011. In this process, we go through the

[6] It is important to note that Figure 1 shows *relative* income levels, expressed in US dollars, i.e., taking the United States as the base country. Expressing GDP per capita in Nigerian Naira would mean Nigeria's GDP per capita is unchanged between the extrapolation from ICP 2005 and ICP 2011, while the other countries show lower GDP per capita levels, since the difference (in US dollars terms) for Nigeria was the largest in relative terms; see online Appendix Figure 1.

[7] This is standard practice in the World Bank's World Development Indicators, the Maddison Project Database (Bolt and van Zanden 2014), and for parts of the Penn World Table (Feenstra, Inklaar, and Timmer 2015).

[8] Though note that the theoretical challenges (Van Veelen 2002) and practical problems (Deaton and Heston 2010; McCarthy 2013b) of relative price measurement will always be a source of differences between different benchmarks.

[9] See World Bank (2015) for a list of major improvements in ICP 2011 methodology compared to ICP 2005.

[10] See, in particular, World Bank (2013).

compilation of purchasing power parities (PPPs) as detailed in World Bank (2008) and World Bank (2015), in particular the linking of regions. In presenting our results, we focus on the implications for relative price levels of GDP and household consumption, which are defined as the PPP for GDP or household consumption, divided by the market exchange rate. We make two sets of adjustments to the original ICP 2005 comparison, namely harmonizing the PPP compilation methodology and correcting the ICP 2005 price data for linking biases. We compare relative price levels extrapolated from our counterfactual ICP 2005 to the results from ICP 2011, analogous to the comparison of relative income levels in Figure 1, focusing on the mean difference between those sets of relative prices and the degree to which those differences vary systematically with income and consumption levels. We find that the methodological harmonization and correction for linking biases reduce the mean difference and that the remaining differences are no longer systematically related to income and consumption levels.

These results have important implications. Most importantly, the difference between the original ICP 2005 and ICP 2011 suggested that datasets based on extrapolating relative prices from a benchmark comparison to earlier years were strongly biased. This, in turn, could have had serious consequences for research based on such datasets, which include the World Development Indicators of the World Bank, the Maddison database (Bolt and van Zanden 2014), and parts of the Penn World Table (Feenstra, Inklaar, and Timmer 2015). Our results suggest that such a fundamental reconsideration is not necessary, since correcting for measurement differences through our counterfactual for ICP 2005 eliminates the systematic differences between extrapolations from the counterfactual for 2005 and the ICP 2011 benchmark comparisons. Furthermore, one concrete outcome of our analysis is two sets of price comparisons, the counterfactual for ICP 2005 and ICP 2011, that are based on the same methods and comparable data. This means they can be fruitfully used in combination, such as in the part of the Penn World Table that presents relative price and income levels based on all available ICP benchmarks. Finally, from a statistical point of view, we have demonstrated the usefulness of backward revisions to relative prices, which is similar in spirit to the backward revision of national accounts time series when introducing new methods.

I. Basic Framework and Initial Comparison

The basic aim of the ICP is to measure the PPP in a country j relative to any other country k. Implicit in this aim is the idea that the law of one price does not hold, because if it did hold we would only need to observe market exchange rates to put GDP or consumption on a comparable basis (abstracting from speculative movements in currencies). The presence of non-traded goods in an economy can be one reason for the law of one price to fail, but see also Rogoff (1996) for other reasons. As a result, a procedure is needed to estimate the economy-wide price level using data on prices and expenditures for individual products i.[11]

[11] We follow the argumentation of Deaton and Heston (2010) that an achievable aim is to measure an index of the price level, though see also Neary (2004) for steps toward a welfare-based "cost-of-living" comparison.

If we choose a so-called superlative index for this purpose, we know from Diewert (1976) that such an index will provide a second-order local approximation to arbitrary preferences. As an (expositional) example, assume the superlative Törnqvist index. The relative price level, or PPP, of country j relative to k at time t can then be written as

$$\log\left(PPP_{jk}^{t}\right) = \frac{1}{2}\sum_{i}\left(s_{ij}^{t} + s_{ik}^{t}\right)\log\left(p_{ij}^{t}/p_{ik}^{t}\right), \tag{1}$$

where $s_{ij} = v_{ij}/\sum_i v_{ij}$, the expenditure shares of product i, and superscript t represents the time period or the year. Note that the relative price level between j and k depends on expenditure shares in both countries.[12] How will this relative price index evolve over time? The standard approach, used in, e.g., the World Bank's *World Development Indicators*, is to assume the change in the relative price level is equal to relative inflation. To extrapolate PPPs from 2005 to 2011, we follow the same approach.[13]

For ICP 2011, we use the results as published in the Summary Report (World Bank 2014) and specifically the purchasing power parities (PPPs) and expenditure levels for GDP and final household consumption expenditure ("consumption"). For ICP 2005, we use the PPPs as originally published in World Bank (2008). The dataset for comparative analysis is constrained mostly by ICP 2005, which covered 146 countries, versus the 177 countries covered comprehensively in ICP 2011. Argentina, Syria, and Lebanon participated in ICP 2005 but not in ICP 2011, and Zimbabwe participated in ICP 2005 but the coinciding period of hyperinflation led to unreliable results. This leaves us with a dataset of 142 countries, which is used in the remainder of the paper. We use GDP deflators to extrapolate GDP PPPs and we use the consumer price index (CPI) to extrapolate consumption PPPs.[14]

Let PPP_{Yjt}^{τ} be the PPP for GDP (Y) in country j at time t based on a benchmark year t[15] and let P_{Yjt} be the GDP deflator in year t in country j. Whenever $t \neq \tau$, the PPP is extrapolated from the benchmark year as

$$PPP_{Yjt}^{\tau} = PPP_{Yj\tau}^{\tau} \times \frac{P_{Yjt}/P_{Yj\tau}}{P_{Ykt}/P_{Yk\tau}}, \tag{2}$$

[12] In the multilateral indexes used in ICP, the price level between j and k also depends (to a lesser degree) on expenditure shares in all countries, see Diewert (2013b).

[13] This is an imperfect approach, both on conceptual grounds—because PPPs should be based on two sets of expenditure shares (see Deaton and Aten 2017)—and practical grounds (McCarthy 2013b), but based on these grounds, we do not expect a systematically different effect in low-income countries compared to high-income countries.

[14] An alternative would be the implicit price deflator of household consumption expenditure from the National Accounts, but results are very similar.

[15] We drop the reference country k here to simplify notation. The reader is reminded that the PPP for any country j is always defined with reference to a base or numeraire country, denoted by k.

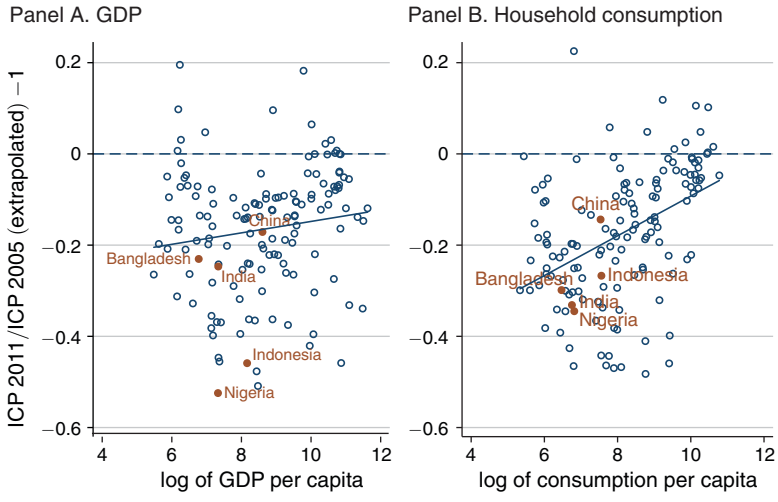

FIGURE 2. DIFFERENCES BETWEEN EXTRAPOLATING ICP 2005 AND ICP 2011

Notes: GDP and consumption per capita are in exchange-rate converted US dollars. Extrapolation from ICP 2005 is done using GDP/consumption inflation rate of each country relative to the United States using equation (2).

Source: Computations based on World Bank (2008, 2015) and World Development Indicators

where k refers to a base or reference country—usually the United States. We can then compare the PPP extrapolated from ICP 2005 with the PPP based on the new ICP 2011 benchmark. We use the following relative difference measure:[16]

$$d_{Yj} = \frac{PPP^{2011}_{Yj2011}}{PPP^{2005}_{Yj2011}} - 1. \tag{3}$$

Figure 2 shows a plot of d_{Yj} against the log of GDP per capita in 2011 (in US dollars) converted using the exchange rate E, $\log(y_j/E_j)$, and d_{Cj} against the log of consumption per capita, $\log(c_j/E_j)$. The GDP and consumption per capita numbers are those reported in World Bank (2015).

The figure illustrates the very large differences between the PPPs that were extrapolated from ICP 2005 and the ICP 2011 PPPs. The countries in Figure 1 had higher income levels based on ICP 2011 than on extrapolations from ICP 2005, which corresponds to lower relative price levels, as shown in Figure 2. For many countries, the ICP 2011 PPPs are between 20 and 40 percent lower, and some countries show even larger differences. Furthermore, there is a positive relationship between income or consumption levels and differences between extrapolated and observed PPP, and this effect is more pronounced for consumption.

To summarize the patterns in Figure 2 and compare these across alternative PPP series in the remainder of this paper, we use three statistics based on the difference measure in (3), namely (for $i = Y, C$ and the set of J countries) the mean difference,

[16] Given that the differences are not small, we use this difference measure rather than the log difference.

Table 1—Differences between PPPs from ICP 2011 and Extrapolation from ICP 2005—Summary Statistics

	All countries	Non-oil countries	Developing economies
GDP			
Mean difference	−0.165	−0.145	−0.192
Root mean squared difference	0.216	0.192	0.236
Coefficient on log(expenditure/capita)	0.013	0.021	−0.018
	(0.007)	(0.007)	(0.012)
Consumption			
Mean difference	−0.176	−0.157	−0.214
Root mean squared difference	0.227	0.206	0.250
Coefficient on log(expenditure/capita)	0.044	0.045	0.018
	(0.007)	(0.006)	(0.012)

Notes: Summary statistics are based on differences in Figure 2 for 142 countries; see the main text for computations. Countries where fuel exports contribute more than 10 percent of real GDP are labeled as "oil countries" (source: Penn World Table, version 8.1). Developing economies are the set of low-income and middle-income countries as defined by the World Bank, so with a GNI/capita level less than $12,746. Robust standard errors of the regression coefficients are shown in parentheses below the coefficients.

Source: Computations based on World Bank (2008, 2015) and World Development Indicators

$\overline{d}_i = \frac{1}{J}\sum_j d_{ij}$, the root mean squared difference (RMSD), $\sqrt{\frac{1}{J}\sum_j d_{ij}^2}$, and the slope coefficient of the regression: $d_{ij} = a_i + b_i \log\left(e_{ij}/E_j\right) + u_{ij}$, where e_{ij} is either GDP or consumption per capita in national currency units and E_j is the exchange rate. These summary measures are best understood by viewing the problem as a forecasting exercise. Until the release of the benchmark ICP 2011 PPPs, the typical approach was to forecast PPPs for 2011 by extrapolating ICP 2005 PPPs using equation (2).[17] Statistic $\overline{d}_i$ is then a measure of the forecasting bias, RMSD is a measure of forecasting uncertainty,[18] and b_i an indication of whether the bias varies systematically with GDP or consumption per capita.[19]

Table 1 shows summary statistics corresponding to the two panels in Figure 2. The average country has a relative price level from ICP 2011 that is about 17 percent lower than based on extrapolations from ICP 2005. This means that the average country is 24 percent richer relative to the US than estimated previously, implying a sizeable narrowing of cross-country income inequality.[20] The root mean squared difference is also comparable for GDP and consumption, at approximately

[17] An exception was Ravallion (2013, 2014), who has argued for a so-called "dynamic Penn effect" based on exchange rate and GDP movements. However, Inklaar (2013) showed that the forecasting performance of the dynamic Penn effect is inferior to inflation-based extrapolation.

[18] In the terminology of the forecasting literature, this would be the root mean squared prediction error, RMSPE, see West (2006).

[19] Note that the average difference and root mean squared difference are not independent of the numeraire country; see also the discussion following Table 1, below. Since typical comparisons are almost exclusively made with the United States as the numeraire this is not a major drawback, but see Diewert (2009) for alternative measures.

[20] Since differences are not small, our difference measure is not taken in log terms, see equation (3) and footnote 16.

22 percent. The main difference between the GDP and consumption differences is that the coefficient on expenditure per capita is much larger for consumption than for GDP, though both are significantly positive. The differences tend to be smaller for non-oil-producing countries, and the mean difference is larger for developing economies than for the sample as a whole.

The "developing economies" column shows the difference in relative prices relative to the United States for the set of developing economies. For the purpose of global poverty analysis, the more relevant relative prices would be comparing only developing economies with, for example, India rather than the United States as the base country; see e.g., Deaton and Dupriez (2011). Computing differences using India as the base country would show a notably smaller mean difference for developing economies, at 5 percent for relative prices of GDP and 12 percent for consumption; the RMSD is 15 percent and 17 percent.[21] The coefficient on the log of expenditure per capita is not affected by the choice of base country and neither are our later results on international income inequality.

In the remainder of this paper, the results in Table 1 will serve as a baseline for comparison with statistics based on counterfactual versions of ICP 2005 that harmonize methodology and resolve linking biases. Ideally, those adjustments will lead to a closer alignment between extrapolations for 2011 based on the counterfactual ICP 2005 PPPs and the actual PPPs from ICP 2011.

II. Methodological Harmonization

The International Comparison Program (ICP) is a major international statistical initiative in which prices and national accounts data are collected from a large number of countries and processed to compile PPPs and internationally comparable real expenditures. The early phases of the ICP were essentially world comparisons where data from all participating countries were treated as a single set. A major drawback of such world comparisons is that it proved hard to collect price data for the same products across such a wide range of countries, and the requirements for global comparability stood in the way of comparability between the more similar countries in each region. A start was made to move away from a single world comparison with ICP 1993, but ICP 2005 was the first comparison with a fully implemented regionalized approach.

The regionalized approach involves compilation of PPPs in two stages, first a regional and then a global comparison. In the regional stage, countries collect and compile data according to regional product specifications, leading to more reliable relative prices within each region. In the second stage, the regional comparisons are linked to complete the global comparison. The methodology for linking regional comparisons was pioneered while data collection for ICP 2005 was already in progress. So it was perhaps not surprising that after the release of ICP 2005, an assessment of the methods used in ICP 2005 for linking regional comparisons identified several deficiencies. These deficiencies relate to the methods for linking prices at the

[21] Though we do not systematically report summary statistics using other base countries, such figures can be derived based on the online dataset accompanying this paper.

level of product categories—referred to as "basic headings" in ICP—and for linking at the aggregate level; see Rao (2013) and Diewert (2013a). To remedy the signaled deficiencies, several methodological innovations were implemented in ICP 2011.

Major changes in methodology suggest that the results from ICP 2005 and ICP 2011 cannot be directly compared. In this paper, we offer a solution to this problem by constructing a counterfactual for ICP 2005 based on ICP 2011 methodology. We briefly describe the major methodological innovations in the ICP 2011—with further details available in World Bank (2015). We then construct counterfactual PPPs for 2005, based on data used in the ICP 2005 computations. The basic data used are the PPPs and national expenditure data at the product category level for all the 146 participating countries. In addition, productivity adjustment data from ICP 2005; information on dwellings; the 2005 ring product list and prices; the 2011 global core list products and their prices in all the participating countries have been obtained from the ICP Global Office at the World Bank.[22]

A. *Methodological Adjustments*

In ICP 2005, the African region employed a transitive and additively consistent aggregation method, the Iklé-Dikhanov-Balk (IDB) method, to compute aggregate PPPs within Africa, while all other regions used the Gini-Elteto-Koves-Szulc (GEKS) method.[23] Bringing all the regions into line, the GEKS method was implemented in all the regions in ICP 2011. The first step in the construction of the counterfactual is thus the replacement of the 2005 PPPs within the African region based on IDB method with PPPs computed using the GEKS method.

Government consumption, covering expenditure on public administration, health, and education, has always been considered "comparison-resistant" in ICP, since market prices for the output of most of these services cannot be observed. So instead of relative output prices, ICP relies on comparing input prices, specifically wages and salaries. A drawback of using input prices is that productivity differences are not taken into account, which means that in lower-income countries, a simple measure of relative wages would understate relative output prices and overstate real expenditure in this category.

In ICP 2005, the Asia-Pacific and African regions recognized the need to make productivity adjustments to relative wage data in their regional comparisons, but no other regions implemented productivity adjustments. Furthermore, in linking the regions, no adjustment was made for productivity differences, thereby leading to an underestimation of the relative price of government consumption in countries from low-income regions relative to countries from high-income regions.

In ICP 2011, a full-fledged productivity adjustment for all the participating economies was implemented. As detailed in World Bank (2015), the ICP 2011 productivity adjustment is based on economy-wide differences in the level of capital per worker. The assumption is that productivity in public administration, health, and education is lower (only) when capital per worker is lower. Under this assumption,

[22] These data are routinely provided to researchers upon request to the ICP Global Office.
[23] See Diewert (2013b) for detailed descriptions of the IDB, GEKS, and other aggregation methods.

relative wages in a country with a low level of capital per worker should be increased by the difference in contribution from capital intensity to labor productivity[24] compared with a country with a high level of capital per worker; see World Bank (2015) for details. To follow the approach used in ICP 2011, we first undo the within-region productivity adjustment in the Asia-Pacific and African regions and then implement the ICP 2011-style productivity adjustment.[25] While this is an important step in harmonizing methodologies between our counterfactual ICP 2005 relative prices and ICP 2011, it will not be able to account for the patterns observed in Figure 2 and Table 1. This is because the productivity adjustment increases government consumption prices in lower-income regions compared to the higher income regions for the counterfactual ICP 2005 prices, thereby increasing the differences relative to ICP 2011.

To compare relative rents of dwellings across regions, the procedures followed in ICP 2005 and ICP 2011 were different from methods used for other types of consumption, see Heston (2013) and World Bank (2015).[26] Ideally, information on the actual (or imputed) rent paid on different types of dwellings would be used for this purpose but such data were not available for all countries; only data on the number of dwellings and some quality characteristics were widely available.

In ICP 2005, the available rental price data was used to link the relative rents for dwellings across regions, see Heston (2013). For ICP 2011, a similar approach was tried but the rental data supplied was judged to be of insufficient quality. Instead a method based on quantity indicators was applied. First, the number of dwellings per capita is used to construct a regional quantity index for dwellings, which is then adjusted for quality differences using a quality index computed as the arithmetic average share of dwellings with electricity, inside water, and private toilet.

To harmonize this aspect of the methodology, we apply the ICP 2011 methodology on dwellings quantity and quality indicators data for 2005 supplied by the ICP Global Office at the World Bank. We combine the per capita volume index with a per capita value index, computed as the ICP 2005 expenditure in the "Actual and imputed rentals for housing" category, converted to US dollars using the current exchange rate and divided by population.[27] This gives a price-level index for each region. The steps involved in implementing the ICP 2011 linking methodology for dwellings on data from 2005 are shown in online Appendix Table A1. In Africa, the Asia-Pacific, and Latin America, relative prices according to the ICP 2011 methodology are lower than under the ICP 2005 methodology, which can thus reduce some of the differences observed in Figure 2 and Table 1.

[24] If the aggregate economy is characterised by a Cobb-Douglas production function with inputs of capital K, labor L, and output elasticity of capital α, this contribution is equal to $(K/L)^{\alpha}$.

[25] See World Bank (2015) for details on the methodology and the Penn World Table, version 8.1 for data on comparative capital stocks and the share of labor income in GDP.

[26] We are grateful to Alan Heston who provided details of the approach used in 2005; Nada Hamadeh for providing data used in 2005 as well as in 2011; and to Paul Konijn for providing us with a copy of the document "*Linking the regions: the case of housing*" outlining the methodology used in linking dwellings data in ICP 2011.

[27] Note that the ICP 2005 data used for the value index also had important shortcomings, with countries reporting implausibly low (imputed) spending in this category—see Deaton (2010). To the extent that such shortcomings may be less in ICP 2011, this could lead to a further difference between ICP 2005 and ICP 2011.

In the implementation of the ICP, the very last step involves linking of regions at higher levels of aggregation, such as household consumption and GDP. Although relative prices for product categories have already been linked across regions, there is a methodological challenge due to the additional requirement of *fixity*. Under this requirement, the comparison of aggregate prices *within* each region should not change as prices are compared across regions. So, for example, the GDP PPP for India relative to China (both in the Asia-Pacific region) should remain fixed when comparing India and China to the United States. The fixity requirement rules out standard aggregation methods.

Diewert (2013a) summarizes a range of methods for global comparisons that maintain fixity. The method used in ICP 2005 treats each region like a super-country,[28] using the country relative price and expenditure data to derive a regional quantity and relative price vector, in terms of a reference country in each region. These regional quantity and price vectors are used to establish the aggregate relative prices between regions. However, it was later found, by Sergeev (2005), that this method was not invariant to the choice of the reference country in each region. ICP 2011, therefore, *switched to the country aggregation and redistribution* (CAR) method, described originally in United Nations (1986). Under this method, aggregate PPPs are first computed using data for all countries in the world, ignoring the fixity requirement. Those aggregate PPPs are used to compute PPP-converted, or real, GDP levels for each country. By summing real GDP across countries, the share in world real GDP of each region can be computed. Finally, the world real GDP share of each region is distributed over the countries in the region using the real GDP shares that were computed based on the within-region PPPs. In constructing the counterfactual comparisons for 2005, we implement the CAR approach instead of the super-country approach.

B. *Counterfactual PPPs Based on Harmonized Methodology*

The first step to constructing a counterfactual for ICP 2005 relative prices is to harmonize the methodology with that used in ICP 2011. In Table 2, we show summary statistics based on the original ICP 2005, from Table 1, and statistics based on a step-wise methodological harmonization. Comparing the first and last column shows that harmonizing the methodology results in smaller mean differences and a lower RMSD for both GDP and consumption. The mean difference and RMSD for the fully harmonized relative prices remain substantial, though, and the coefficients on expenditure per capita are significant and positive.

Going through the various stages of the harmonization, the most substantial change results from harmonizing the within-Africa PPPs, while implementing the CAR method, productivity adjustment and dwellings adjustment matter relatively less. As predicted, the productivity adjustment increases mean differences and the coefficient on expenditure per capita for GDP. In addition to the methodological innovations in 2011 discussed here, ICP 2011 also introduced a new method for

[28] The term is from Deaton and Aten (2017).

Table 2—Differences between ICP 2011 and Extrapolations from ICP 2005—Counterfactuals Based on Harmonized Methodology

	Original	CAR method	+ Within-Africa harmonization	+ Productivity adjustment	+ Dwellings adjustment
GDP					
Mean difference	−0.165	−0.154	−0.124	−0.157	−0.151
Root mean squared difference	0.216	0.213	0.187	0.210	0.206
Coefficient on log(expenditure/capita)	0.013 (0.007)	0.023 (0.007)	0.013 (0.007)	0.024 (0.007)	0.023 (0.007)
Consumption					
Mean difference	−0.176	−0.143	−0.128	−0.128	−0.119
Root mean squared difference	0.227	0.276	0.204	0.204	0.198
Coefficient on log(expenditure/capita)	0.044 (0.007)	0.060 (0.010)	0.050 (0.007)	0.050 (0.007)	0.048 (0.007)

Notes: The column labeled "Original" is from Table 1. "CAR method" uses the regional linking methodology of ICP 2011, "+ Within-Africa harmonization" uses the GEKS method rather than the IDB method to compare prices within the African region; "+ Productivity adjustment" also implements the productivity adjustment for public wages for the Eurostat/OECD region and Latin America, and "+ Dwellings adjustment" also implements the ICP 2011 method for comparing the (rental) price of dwellings across regions. This column corresponds to a set of PPPs where the methodology has been harmonized to the extent possible. Robust standard error of the regression coefficients are shown in parentheses below the coefficients.

Source: Computations are based on World Bank (2008, 2015), World Development Indicators, ICP 2005: Data for Researchers, PWT 8.1 (Feenstra, Inklaar, and Timmer 2015), and data on dwellings provided by the ICP Global Office.

measuring relative prices in construction. ICP 2005 used the "basket of construction components" (BOCC) method—see McCarthy (2013a) for details—but it was replaced in ICP 2011 by a simpler method based on the prices of basic construction materials, different types of labor, and the hire of machinery and equipment. The information required to implement the ICP 2011 construction methodology was not available from the ICP 2005 data sources.

III. Accounting for Linking Bias in ICP 2005

This section aims to establish whether the linking of regions at the level of product categories led to biases in ICP 2005, motivated by the fact that the linking procedure represents a major difference between ICP 2005 and ICP 2011. In ICP 2005, linking was based on price data for only 18 countries, the so-called ring countries, which were deemed to be representative for the full set of countries in their respective regions. [29] The ring countries collected prices for items on the global ring product list and those prices were then used to link PPPs from different regions; see also Deaton and Heston (2010) and Vogel (2013).[30] In contrast, the linking process in ICP 2011 was based on prices collected by all 177 participating countries for

[29] The ring countries are Cameroon, Egypt, Kenya, Senegal, South Africa, and Zambia in Africa; Hong Kong, Malaysia, Philippines, and Sri Lanka in Asia-Pacific; Brazil and Chile in Latin America; Jordan and Oman in Western Asia; and Estonia, Japan, Slovenia, and the United Kingdom in Eurostat/OECD.

[30] See Rao (2013) on the linking methodology at the product category level.

products on a global core list (GCL). As a result, World Bank (2015) emphasizes changes to linking methodology as a major innovation in ICP 2011.

We consider two possible types of linking bias in ICP 2005 compared with ICP 2011. The first is whether the particular selection of 18 ring countries would induce a bias compared to the use of data for all the participating countries in the linking process. We term this bias *ring country selection bias*. The second issue relates to the items included in the ring product list. A close examination of the products included in the ring list and the product specifications used for pricing purposes revealed a strong rich country bias in the products (Deaton 2010). The list included items like Uncle Toby oats, a bottle of Heineken beer, Bordeaux wine, and a Peugeot 408, all with precisely specified characteristics, such as climate control on the Peugeot 408. This is suggestive of a rich country bias in the ring list product specifications. Such bias in the product specifications is not necessarily problematic, but it does require more careful judgment in lower income countries, where products that are unrepresentative should not be priced. It seems likely that the required level of judgment was not exercised, as evidence presented by Deaton (2010) suggests that the use of the ring product list led to relatively high price levels in low-income ring countries. We refer to this as *ring product selection bias*.

A. *Ring Country Selection Bias*

We cannot use data for 2005 to establish whether there was any bias from the specific selection of ring countries. However, given the price data for all particpating countries in 2011, we can examine if using the particular set of 18 ring countries selected in 2005 would have introduced any systematic bias in the (2011) linking factors used in linking relative prices for product categories.[31] The assumption here is that the ring country bias we find in 2011 would have been present to the same degree in 2005 as well.

To estimate this bias, we rely on a version of the country-product-dummy (CPD) model, a model that is widely used in the ICP.[32] We specify a simple regression model that explains log prices P of product i in country j located in region r using dummies for each product, D_i; each region, D_r; a "ring country" dummy R (which takes value 1 if the country is one of the 18 ring countries); and interactions between the regional dummies and the ring country dummy:

$$\log P_{ijr} = \sum_i \beta_i D_i + \sum_r \gamma_r D_r + \eta R + \sum_r \delta_r (D_r \times R) + \epsilon_{ijr}. \tag{4}$$

In this regression model, the dependent variables are the prices of items in different ring countries which are are first converted into regional numeraire currencies using relative price measures for product categories within the different regions. For

[31] The lowest level of disaggregation at which expenditure data and weights are available is referred to in ICP as the "*basic heading*" (see Rao 2013 for details), we will refer to this level as product categories. Relative prices for product categories are an aggregate of item-level level prices surveyed in different countries.

[32] The CPD model was originally proposed by Summers (1973), where it is shown to be an effective device to fill gaps in price data. Details of the CPD model and its use in the ICP can be found in Rao (2013).

Table 3—Estimating the Difference in 2011 Pricing of Global Core List Products in Ring Countries versus Other Countries

Region	Coefficient γ_r	Region × ring country	Coefficient δ_r
Africa	2.119	Africa × ring country	0.003
	(0.009)		(0.022)
Asia-Pacific	2.140	Asia-Pacific × ring country	0.019
	(0.013)		(0.037)
Latin America	6.446	Latin America × ring country	0.019
	(0.011)		(0.026)
Western Asia	−1.133	Western Asia × ring country	−0.022
	(0.018)		(0.042)
	Coefficient η		
Ring country dummy	−0.006		
	(0.014)		

Notes: The table shows regression results where the log of prices of 56,011 consumption items for 2011, converted to the regional base country's currency, are explained by dummies for each region, whether a country was a ring country in ICP 2005, interactions between this ring country dummy and the regional dummies, and dummies for each of the 1,192 items (not shown); see also equation (4). Items are weighted by their importance and an estimate of the share of consumption expenditure accounted for by the item; see equation (5). The Eurostat/OECD region is taken as the base region and hence the coefficients should be interpreted as price differences relative to the Eurostat/OECD. Robust standard errors, clustered by product, of the regression coefficients are shown in parentheses below the coefficients.

Source: ICP 2011: Data for Researchers, national average item prices

example, prices of different varieties of rice collected from countries in the Asia-Pacific region are converted into Hong Kong dollars using relative prices for the "rice" product category. These converted prices are used in (4). This model, without the ring country dummy and interaction terms, is used in ICP to derive the linking factors, see Vogel (2013). However, to us the δ_r parameters are of main interest since a significant coefficient would indicate that pricing in the ring countries in that region are significantly different from pricing in the other countries in the region.

Table 3 reports the results of a pooled regression using data for detailed commodity groups, based on 56,011 price observations.[33] At this level of detail, we do not know how much is spent on each product, only whether it was considered an important (i.e., representative) product or not, with important products receiving a weight of three and other products a weight of one in ICP. However, we do know expenditure at the level of product categories and this is important to take into account, as the effect of a bias in product prices on GDP and consumption PPPs will depend on the relative importance of that product category in overall consumption. Since the number of individual products that are priced within product categories vary substantially (between 1 and 42 products), we construct a composite weight as

$$w_{ijr} = I_{ijr} \times \frac{1}{n_{jr}^{k}} \times \frac{v_{jr}^{k}}{\sum_{k} v_{jr}^{k}}, \tag{5}$$

[33] The regressions includes data from the 148 countries that participated in ICP 2011 and that were part of the regions shown in Table 3. If these countries had priced all the 692 GCL products, we would have 102,416 observations, so not all products were priced in all countries.

where I are the ICP 2011 importance weights, superscript k identifies the product category to which product i belongs, n_{jr}^{k} is the number of products that are priced, and v_{jr}^{k} is expenditure, both for product category k in country j, part of region r.

The regional coefficients in the left panel provide the average regional linking factors γ_r. For example, the table shows that 8.50 Hong Kong dollars (exponential of 2.140, the coefficient for the Asia-Pacific) for one US dollar is the linking factor when all the countries are used in linking the regions. We also observe that the coefficient of ring country dummy (η) is numerically very small and statistically insignificant. This means that the selection of ring countries on average has no influence on the linking factors. Most importantly, we find that none of the δ_r parameters is significantly different from zero. This indicates that, at least in 2011, the pricing patterns in the ICP 2005 ring countries were not systematically different from that in the non-ring countries. There is thus no evidence to suggest that the choice of this particular set of 18 countries in ICP 2005 for regional linking led to a bias in the overall PPPs.

B. *Ring Product Selection Bias*

In ICP 2005, the 18 ring countries priced items from the ring product list and ideally, this list should contain products that are part of a representative consumption bundle in all countries.[34] In practice, this is much harder, as discussed in more detail by Deaton and Heston (2010). For instance, a bottle of Heineken beer was on the ring product list, but if this is a premium brand in poor countries and a standard choice of beer in rich countries, beer prices in the poor country will be relatively higher than would be the case for beer in general. Deaton (2010) presented suggestive evidence that points to a "rich country" bias in the ring product list.

Given this potential for bias, a *global core list* (GCL) of products was developed for ICP 2011 where more care was taken to ensure that goods and services are representative in all regions of the world. Moreover, clearer instructions were given to price collectors that products should not be priced if they are not representative in a particular country. The ICP 2011 GCL product approach seems to have been fairly successful as the GCL had significant overlap with the regional product lists. Of the 692 consumption products on the GCL, 610 matched with the regional list in Africa, 412 with Asia, 394 with OECD-Eurostat, 489 with Latin America and the Caribbean, and 606 with the list used in Western Asia.

To explicitly test whether the product selection in drawing up the ring product list was a source of bias, we exploit the fact that the ICP 2005 ring countries participated in two sets of price comparisons: one based on the regional product list as part of the within-region comparisons and the other based on the ring product list for the purpose of computing cross-region linking factors. It is important that there are multiple ring countries from each region, as it allows us to test whether price comparisons based on each regional product list and the global ring product list lead to similar

[34] Our analysis focuses on household consumption products, where the regional product lists differed markedly from global product lists in ICP 2005 and 2011. For government consumption and investment, the global lists were adopted with only minor variations across regions.

price relatives between, say, Hong Kong and Malaysia (both in the Asia-Pacific region) or between Cameroon and South Africa (both in the African region).

The regional lists were drawn up to be representative of products within each region, so it seems reasonable to assume that the results based on the regional price comparison are a more accurate measure of relative prices between the ring countries within each region than the relative price differences implied by the ring product list. Indeed, in the compilation of ICP 2005 PPPs, the ring product list prices are only used for linking prices between regions, not for price comparisons within the region. So if we find that the relative prices within each region based on the ring product list prices differ systematically from the relative prices implied by the regional product prices, we take this as evidence of a bias from the particular selection of the products on the ring product list. Note that these bias estimates rely only on within-region prices and that the bias to across-region linking factors could be larger (or smaller) than what we find here; we return to this issue below.

We use a CPD model and apply it to data for the ICP 2005 ring countries to test for the presence of product selection bias:

$$\log P_{ijr} = \sum_i \beta_i D_i + \eta_{jr} D_{jr} + \epsilon_{ijr}. \tag{6}$$

As before, the D_i are dummies for different products; additionally, D_{jr} are dummies for every ring country j in region r. The omitted country dummy is the region's base country, which is chosen as the ring country with the highest income level. Importantly, P is the price of product i *relative to the PPP* for the product category to which product i belongs. PPPs referred to here are for the product categories, computed using regional products lists, and expressed relative to the same base country of each region, as in equation (4).

This normalization allows us to interpret the estimates of η_{jr} as an estimate of ring product selection bias: if η_{jr} is significantly positive, prices of products from the ring product list are high compared to prices from the regional product list in country j, relative to the base country of region r. We choose this base country as the country with the highest income level of the region so that a positive and significant η_{jr} indicates that ring product prices *overstate* prices in the low-income ring country.

Equation (6) is estimated for each region and the results are shown in Table 4. Observations are weighted by the estimated expenditure share for each product, based on equation (5) but omitting the importance weights I that were not available for ICP 2005. The table shows clear evidence of ring product bias in the Asia-Pacific region, where ring product prices are between 8 percent (in Malaysia) and 20 percent (Sri Lanka) higher than would be expected based on the regional price comparison. Significant and large coefficients are also clear in Africa, for Cameroon (+18.9 percent) and Senegal (+8.6 percent). Smaller, though significant, biases are apparent in Slovenia and Brazil. So, compared to the Eurostat/OECD region (of which the global base country, the United States, is part), Table 4 provides evidence that prices in Asia-Pacific and Africa were notably higher than they should have been, and with a smaller bias in the same direction for Latin America.

Table 4—Estimating the Difference between Ring Product and Regional Prices in 2005 between Regional Base Country and 2005 Ring Countries

Region (base country)	Country	Country dummy	Observations	Products
Africa (South Africa)	Cameroon	0.189 (0.038)	2,999	736
	Egypt	–0.038 (0.047)		
	Kenya	0.044 (0.047)		
	Senegal	0.086 (0.046)		
	Zambia	–0.054 (0.053)		
Asia-Pacific (Hong Kong)	Sri Lanka	0.198 (0.054)	2,175	683
	Malaysia	0.080 (0.036)		
	Philippines	0.113 (0.047)		
Eurostat-OECD (UK)	Estonia	0.036 (0.023)	2,031	646
	Japan	0.037 (0.030)		
	Slovenia	0.050 (0.023)		
Latin America (Chile)	Brazil	0.069 (0.032)	594	297
Western Asia (Oman)	Jordan	0.047 (0.050)	710	355

Notes: The table shows regression results based on equation (6), where the log of consumption ring item prices for 2005, converted to the regional base country's currency using product category PPPs, are explained by country dummies and dummies for each of the products (not shown). Items are weighted by the estimated expenditure share for each product, based on equation (5), but omitting the importance weights I that were not available for ICP 2005. Robust standard errors, clustered by product, of the regression coefficients are shown in parentheses below the coefficients.

Source: ICP 2005: Data for Researchers, national average item prices

As a "placebo test" we run the regressions from Table 4 but now using the 2011 GCL price data and the results are presented in Table 5. Given the greater care in devising the product list and in collecting the prices, we would expect little or no evidence of product selection bias and the results conform with our expectations, for the most part. The exception is Malaysia, where GCL prices were significantly *lower* than regional prices. For all other ring countries, though, the coefficients are not significantly different from zero.

C. *Linking Bias Estimates from Deaton and Aten* (2017)

Deaton and Aten (2017) were the first to examine the sources of systematic and major discrepancies between 2011 ICP and extrapolations from the 2005 benchmark.

Table 5—Estimating the Difference in Global Core List Product Prices and Regional Prices in 2011 between Regional Base Country and 2005 Ring Countries

Region (base country)	Country	Country dummy	Observations	Products
Africa (South Africa)	Cameroon	0.018	2,757	561
		(0.037)		
	Egypt	0.026		
		(0.043)		
	Kenya	−0.006		
		(0.033)		
	Senegal	−0.012		
		(0.042)		
	Zambia	0.005		
		(0.039)		
Asia-Pacific (Hong Kong)	Sri Lanka	−0.057	1,387	402
		(0.049)		
	Malaysia	−0.147		
		(0.042)		
	Philippines	−0.044		
		(0.039)		
Eurostat-OECD (UK)	Estonia	−0.026	1,330	380
		(0.025)		
	Japan	−0.042		
		(0.026)		
	Slovenia	0.013		
		(0.021)		
Latin America (Chile)	Brazil	0.000	508	254
		(0.034)		
Western Asia (Oman)	Jordan	0.026	906	453
		(0.036)		

Notes: The table shows regression results based on equation (6), where the log of consumption global core list product prices for 2011, converted to the regional base country's currency using product category PPPs, are explained by country dummies and dummies for each of the products (not shown). Products are weighted according to equation (5). Robust standard errors, clustered by product, of the regression coefficients are shown in parentheses below the coefficients.

Source: ICP 2011: Data for Researchers, national average item prices

Their approach to identifying a linking bias in ICP 2005 differs from ours. They estimated country-level (consumption) PPPs for the ICP 2005 ring countries based on the ICP 2005 ring product list and based on the ICP 2011 GCL. They then compared the change in these PPPs between 2005 and 2011 to the relative inflation rate as in equation (2), taking the United Kingdom from the Eurostat-OECD region as the numeraire. For the high-income ring countries, they find that relative inflation rates approximate the change in PPPs quite well, but for the low-income countries they find smaller changes in PPPs than implied by relative inflation—see Deaton and Aten (2017, table 3).[35] If one assumes that GCL prices are unbiased (which is broadly supported by the results in Table 5), then ICP 2005 relative prices for

[35] This figure from their paper provides a clear separation between high-income ring countries (Brazil, Chile, Estonia, Hong Kong, Japan, Oman, Slovenia, South Africa, and the United Kingdom) and low-income ring

the low-income ring countries were biased upwards. They conclude that the ring product prices may have overstated price levels in the low-income ring countries in Western Asia, Africa, and the Asia-Pacific by approximately 25 percent.

This is a larger adjustment than implied by our analysis in Table 4, as the biases we find do not exceed 20 percent. However, there is reason to believe that the estimates in Table 4 are conservative, since our method can only identify a bias relative to the regional base country, not a bias relative to the global base region, Eurostat-OECD. Especially in Oman (Western Asia) and South Africa (Africa), consumption levels are notably lower than in the United Kingdom, the Eurostat-OECD base country. This could mean that the ring list products may not have been representative in those countries either, which would imply a larger bias than we can identify in our estimations. But while our estimates in Tables 3–5 may not be able to pinpoint the exact magnitude of the linking bias, they do provide clear evidence that such bias was present in ICP 2005 and not in ICP 2011.

IV. Counterfactual PPPs for the Year 2005, ICP 2005C

In addition to the methodological harmonization and the adjustment for linking bias, we also make an adjustment for urban bias in the consumption prices for China. It is well documented that consumption prices provided by China to the Asia-Pacific regional comparison in 2005 were collected from only 11 capital cities and their surrounding areas (see ADB 2007 and Word Bank 2008).[36] The general consensus from Deaton and Heston (2010), Chen and Ravallion (2010a, b), and Feenstra et al. (2013) is that the Chinese consumption prices used in ICP 2005 were biased upward relative to national average prices by 20 percent or more.[37] In contrast, the Chinese price survey for ICP 2011 was much more comprehensive, covering urban and rural areas in each of the country's provinces (see ADB 2014). As a result, it seems likely that for ICP 2011, Chinese price collection did achieve the goal of national average prices rather than urban-biased prices. In constructing the counterfactual for 2005, we make a 20 percent upward adjustment to consumption prices for China.[38]

Accordingly, our preferred counterfactual is constructed on the basis of the adjustments we derive for the methodological harmonization, we use the linking bias estimates from Deaton and Aten (2017), and we include an adjustment for urban bias in China in ICP 2005.[39] We label this ICP 2005C and the rest of the paper focuses on the preferred ICP 2005C counterfactual. In the online Appendix, we present results for a more conservative counterfactual based on the regression coefficients in Tables 3–5.

countries (Cameroon, Egypt, Jordan, Kenya, Malaysia, Philippines, Senegal, Sri Lanka, and Zambia). We follow this delineation.

[36] The Global Office of the ICP at the World Bank and the Asian Development Bank devised a mechanism to derive national average prices for China based on price data from the 11 cities. However, the method failed to account for rural-urban and regional price differentials.

[37] PWT versions 7 and 8 implemented the same adjustment. A similar adjustment was used in Chen and Ravallion (2010a, b) for estimating absolute poverty in China.

[38] This adjustment is alongside the adjustment for linking bias as the size of the urban bias in Chinese consumption prices was determined based on data for China or for other countries in the Asia-Pacific region, so linking bias would not affect those estimates.

[39] In online Appendix Table A2, we show the differences when only accounting for linking bias and urban bias.

Table 6—Differences between ICP 2011 and Extrapolations from ICP 2005C Counterfactual with Methodological and Bias Adjustments

	Original	Counterfactual			
	All countries	All countries	Non-oil countries	Developing economies	All countries, population-weighted
GDP					
Mean difference	−0.165	−0.088	−0.062	−0.100	−0.036
Root mean squared difference	0.216	0.168	0.139	0.179	0.170
Coefficient on log (expenditure/capita)	0.013 (0.007)	−0.001 (0.007)	0.006 (0.007)	−0.033 (0.014)	0.016 (0.011)
Consumption					
Mean difference	−0.176	−0.018	0.002	−0.021	0.052
Root mean squared difference	0.227	0.144	0.126	0.155	0.216
Coefficient on log (expenditure/capita)	0.044 (0.007)	0.000 (0.007)	0.000 (0.007)	−0.023 (0.014)	0.003 (0.025)

Notes: The first column (ICP 2005) is from Table 1. ICP 2005C refers to the counterfactual ICP 2005 prices that harmonize methodology and correct for linking biases. Countries where fuel exports contribute more than 10 percent of real GDP are labeled as "oil countries" (source: Penn World Table, version 8.1). Developing economies are low-income and middle-income countries as defined by the World Bank, so with a GNI/capita level less than $12,746. Robust standard errors of the regression coefficients are shown in parentheses below the coefficients.

Source: Computations are based on World Bank (2008, 2015), World Development Indicators, and computations based on ICP 2005 and ICP 2011: Data for Researchers.

A. *Differences between ICP* 2011 *and Extrapolations from ICP* 2005C

In Table 6, we provide an assessment of the differences between ICP 2011 and ICP 2005C. Extrapolating from ICP 2005C leads to substantially lower differences with ICP 2011, with the mean difference for GDP reduced from −0.165 to −0.088 and for consumption from −0.176 to −0.019. The remaining differences are no longer systematically larger in lower income countries—within the group of developing economies, there even seems to be a reverse pattern. The root mean square differences are also substantially reduced for the counterfactual. Comparing the effect of harmonizing methodology (Table 2) to the figures in Table 6 shows that the linking bias was the main factor behind the large differences between extrapolations from ICP 2005 and ICP 2011 results.

At the GDP level the average differences are still sizable at −8.8 percent, and the RMSD for GDP and consumption imply substantial (idiosyncratic) country-specific differences. There are a number of possible reasons for this. First, we have not been able to implement any adjustment for the new methodology for construction used in 2011. Second, ICP does not survey export and import prices but uses exchange rates instead. This means that differences in terms of trade are not taken into account, despite their importance in cross-country comparisons, see Feenstra, Inklaar, and Timmer (2015). McCarthy (2013b) provides a more extensive set of conceptual and measurement reasons why two subsequent benchmarks may not align.[40] Finally, ICP is a survey of prices so measurement error, reflected in

[40] See also e.g., Dalgaard and Sørensen (2002).

Table 7—GDP per Capita in 2005 in Selected Countries, ICP 2005 versus ICP 2005C

	GDP per capita (in US$)		ICP 2005C/ ICP 2005—1
	ICP 2005	ICP 2005C	
China	4,091	5,875	44%
India	2,126	2,722	28%
Kenya	1,359	1,519	12%
Ethiopia	591	605	2%
Brazil	8,596	8,240	−4%
World GDP ($trillion)	54.8	59.8	10%

Source: World Bank (2008) and authors' computations

e.g., the RMSD, will always be present. Estimates of the standard error of PPPs are on the order of 20 percent—Deaton (2012) and Rao and Hajargasht (2016)—so the RMSD of 14–17 percent in the second column of Table 6 looks reasonable from that perspective. This is no different when estimating relative prices using information on income per capita (i.e., the Balassa–Samuelson effect); the root mean squared error (RMSE) of such regressions is also in the order of 20 percent, despite an R^2 of around 70 percent and highly significant coefficients.[41] These results confirm the need to conduct ICP at regular intervals as extrapolations can only measure relative prices across countries with error.

B. *Implications for Income Levels and International Inequality*

At the time of the release of results from ICP 2005, commentaries from the economists and researchers focused on the large reduction in the size of the world economy and of the Chinese and Indian economies compared to the extrapolations from the 1993 benchmark results, available at that time from the *World Development Indicators* and the *Penn World Table* 6.3; see e.g., Deaton and Heston (2010) and Deaton (2010). It is useful to revisit this issue and compare the differences between ICP 2005 and ICP 2005C.

At the time of release of ICP 2005 (World Bank 2008), the general commentary focused on the downward adjustment to the size of Chinese and Indian economies of 40 and 36 percent, respectively, compared to what was expected at that time based on the *World Development Indicators*. In Table 7, we find the ICP 2005C proposed in this paper pegs back the downward adjustment. In the case of China, the adjustment is 44 percent, bridging the 40 percent gap identified at that time. For India, the counterfactual provides an upward revision of 28 percent of per capita income, which is somewhat lower than the 36 percent identified in 2008. We find smaller upward adjustments to Kenya, Ethiopia, and most of the African and Western Asian countries. In the case of Brazil, the counterfactual implies a downward revision of 4 percent; for most of the rich countries the revisions are marginal. The size of the world economy is 10 percent larger based on ICP 2005C than under ICP 2005,

[41] Based on either ICP 2005 or ICP 2011 data; results are available upon request.

Table 8—GDP per Capita in 2011 for Selected Countries in PPP Converted US Dollars: Extrapolations from Actual ICP 2005 and Counterfactual ICP 2005

	ICP 2011	Extrapolations from ICP 2005		Difference ICP 2005 and ICP 2011	
	Actual	Actual	Counterfactual	Actual	Counterfactual
Kenya	2,136	1,729	1,933	0.24	0.11
Zambia	3,155	1,745	2,047	0.81	0.54
India	4,735	3,566	4,566	0.33	0.04
China	10,057	8,335	11,971	0.21	−0.16
Jordan	11,169	5,844	6,771	0.91	0.65
Brazil	14,639	11,785	11,298	0.24	0.30
United Kingdom	35,091	36,153	37,471	−0.03	−0.06
Germany	40,990	37,729	39,224	0.09	0.05
United States	49,782	49,782	49,782	0.00	0.00

Notes: The column "ICP 2011" shows 2011 GDP per capita as reported in World Bank (2014, table 6.1), based on ICP 2011 PPPs. The column "ICP 2005" shows 2011 GDP per capita based on the actual World Bank (2008) ICP 2005 PPPs or the counterfactual 2005 PPPs developed in this paper, extrapolated to 2011 using the change in GDP deflator in each country relative to the change in the United States. The column "Difference" shows the percentage change of the ICP 2011 column relative to the ICP 2005 columns.

Source: World Bank (2008, 2015), World Development Indictors, and authors' computations

which more broadly counteracts the surprising downward adjustment that was noted with the release of ICP 2005.

In Table 8, we compare the extrapolation from ICP 2005 and ICP 2005C to the ICP 2011 results for a selected set of countries. The final column shows a significant reduction in the differences between the extrapolations and actual ICP 2011 results when the counterfactual is used. However, some differences are still large for countries like Zambia and Jordan. For India the difference reduces to 4 percent, whereas the difference goes down to −16 percent from +21 percent for China. This change is not unexpected as the bias adjustments include a 20 percent downward adjustment to account for the urban bias in Chinese prices in 2005.

The ICP 2011 results showed a significant increase in per capita real income of low-income countries compared to extrapolations from 2005, while there was no appreciable difference for the high-income countries. This suggests that between-country income inequality is lower according to ICP 2011 than based on extrapolations from ICP 2005. We analyze this more formally by examining the trend in the population-weighted Gini coefficient. This provides an indication of what Milanovic (2012) refers to as type-II international income inequality. This measure stops short of a full global inequality measure since it does not take into account inequality within countries, but (in contrast to type-I international income inequality) it does give greater weight to more populous countries.[42]

Table 9 shows that the release of ICP 2011 appeared to result in a significant decline of international income inequality, from a Gini coefficient of 0.526 for GDP per capita based on extrapolations from ICP 2005 to 0.478 based on ICP 2011 PPPs; consumption per capita shows a similar difference. The difference between the Gini

[42] The Gini coefficient is related to the Lorenz curve, which plots cumulative income shares against cumulative population shares when ranking countries by their income/capita level. Specifically, the Gini coefficient is the area between the diagonal and the Lorenz curve divided by the area below the diagonal.

Table 9—Gini Coefficients for GDP and Consumption per Capita in the Year 2011 Based on Alternative PPPs

	GDP	Consumption
ICP 2005	0.527	0.565
ICP 2005C	0.487	0.510
ICP 2011	0.479	0.513

Notes: Gini coefficients are population-weighted and computed based on the 142 countries in the dataset, covering approximately 95 percent of world population. The PPPs are either from the observed ICP 2011 data or extrapolations from ICP 2005 or ICP 2005C. The Lorenz curves corresponding to these Gini coefficients are shown in online Appendix Figure A2.

Source: World Bank (2008, 2015) and authors' computations

coefficients based on ICP 2005C PPPs and ICP 2011 PPP is much smaller: 0.485 versus 0.478. As the corresponding Lorenz curves in online Appendix Figure A2 show, inequality based on ICP 2005C and ICP 2011 is lower (according to the Lorenz criterion) than based on ICP 2005, while—because of crossing Lorenz curves—inequality based on ICP 2011 is not unambiguously higher or lower than based on ICP 2005C.

V. Conclusions

The publication of the ICP 2011 results in April 2014 sparked considerable discussion given the large increases in real GDP per capita of lower income countries. The changes to economic geography, global inequality, and poverty implied by ICP 2011 are significant and warranted an explanation. Deaton and Aten (2017) were the first to analyze this issue in depth and their findings point toward the data and methods used for linking the regions in the ICP 2005 round.

The starting point for our research is the recognition that there have been major innovations in the methodology used in the ICP 2011, which are well documented in the Final Report for ICP 2011; see World Bank (2015). In this paper we have succeeded—to a large extent—in applying the methodology as used for ICP 2011 to the data from ICP 2005. Another important result of our paper is that price data used for linking the different regions of the world were biased in ICP 2005: lower income countries involved in the linking process collected prices for too many products that were nonrepresentative, as judged by the much higher prices we observe relative to the within-region relative prices of these countries. In contrast, we find no evidence of a bias in this direction based on the ICP 2011 price data, which confirms that those later PPPs are more reliable. We construct a counterfactual, called ICP 2005C, based on harmonized methodology and corrected for the effects of linking bias. The main finding is that extrapolations from ICP 2005C differ much less from the actual ICP 2011 PPP and that the remaining differences are no longer significantly larger in lower income countries. Furthermore, the degree of cross-country income inequality is comparable based on (extrapolations from) ICP 2005C and ICP 2011, but unambiguously higher based on (extrapolations from) ICP 2005.

From a statistical point of view our research suggests that constructing counterfactual PPPs is a useful exercise whenever major methodological changes are introduced. Such counterfactuals can reduce the size of the revisions following a new PPP release and highlight how much of the revision is due to the methodological changes and how much due to new measurements in a different year. Our results also make a case for more frequent price comparisons. Since a price comparison in any year will provide results that are measured with a sizeable degree of error, more frequent measurement will prevent outliers in any one year from distorting international comparisons of living standards for a substantial period of time. Furthermore, the introduction of new methodologies can more easily be phased in gradually if there is an ongoing statistical infrastructure for surveying prices, again leading to fewer large revisions.

From a researcher point of view, our results can be seen as broadly comforting. Many datasets used for cross-country analysis of economic performance rely on a single set of PPPs to construct relative income levels, which are then extrapolated to other years using national growth rates. Major differences across PPP sets that are systematically larger (or smaller) in low-income countries would argue against this practice and call into question research based on such datasets, which include the World Development Indicators, the Maddison data, and (parts of) the new Penn World Table. Our counterfactual for 2005 suggests that such revisionism is not necessary and that the standard extrapolation method is not systematically biased. At the same time, the large (idiosyncratic) differences between the two benchmarks for individual countries suggests that relying on a single benchmark for data on individual countries is hazardous. Our counterfactual for 2005 provides a helpful and consistent second point of reference in addition to the new data for 2011.

REFERENCES

Aghion, Philippe, Ufuk Akcigit and Peter Howitt. 2014. "What Do We Learn From Schumpeterian Growth Theory?" In *Handbook of Economic Growth*, Vol. 2, edited by Philippe Aghion and Steven N. Durlauf, 515–63. Amsterdam: North-Holland.

Almås, Ingvild. 2012. "International income inequality: measuring PPP bias by estimating Engel curves for food." *American Economic Review* 102 (2): 1093–1117.

Anand, Sudhir, and Paul Segal. 2008. "What Do We Know about Global Income Inequality?" *Journal of Economic Literature* 46 (1): 57–94.

Asian Development Bank (ADB). 2007. *2005 International Comparison Program in Asia and the Pacific: Purchasing Power Parities and Real Expenditures*. Manila, Philippines: Asian Development Bank.

Asian Development Bank (ADB). 2014. *Purchasing Power Parities and Real Expenditures: Summary of Results and Findings of the 2011 International Comparison Program*. Manila, Philippines: Asian Development Bank.

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

Bolt, Jutta, and Jan Luiten van Zanden. 2014. "The Maddison Project: collaborative research on historical national accounts." *Economic History Review* 67 (3): 627–51.

Chen, Shaohua, and Martin Ravallion. 2008. "The Developing World Is Poorer Than We Thought, But No Less Successful In The Fight Against Poverty." World Bank Policy Research Working Paper 4703.

Chen, Shaohua, and Martin Ravallion. 2010a. "China Is Poorer Than We Thought, But No Less Successful In The Fight Against Poverty." World Bank Policy Research Working Paper 4621.

Chen, Shaohua, and Martin Ravallion. 2010b. "The Developing World Is Poorer Than We Thought, But No Less Successful in the Fight Against Poverty." *Quarterly Journal of Economics* 125 (4): 1577–1625.

Ciccone, Antonio, and Marek Jarociński. 2010. "Determinants of Economic Growth: Will Data Tell?" *American Economic Journal: Macroeconomics* 2 (4): 222–46.

Dalgaard, Esben, and Henrik Sørensen. 2002. "Consistency between PPP Benchmarks and National Price and Volume Indices." Presented at the 27th General Conference of the International Association for Research on Income and Wealth, Djurhamn, Sweden, August 18–24.

Deaton, Angus. 2010. "Price Indexes, Inequality, and the Measurement of World Poverty." *American Economic Review* 100 (1): 5–34.

Deaton, Angus. 2012. "Calibrating measurement uncertainty in purchasing power parity exchange rates." Presented at the International Comparison Program (ICP) Technical Advisory Group, Washington, DC, September 17–18.

Deaton, Angus, and Betina Aten. 2017. "Trying to Understand the PPPs in ICP 2011: Why Are the Results So Different?" *American Economic Journal: Macroeconomics* 9 (1): 243–64.

Deaton, Angus, and Olivier Dupriez. 2011. "Purchasing Power Parity Exchange Rates for the Global Poor." *American Economic Journal: Applied Economics* 3 (2): 137–66.

Deaton, Angus, and Alan Heston. 2010. "Understanding PPPs and PPP-Based National Accounts." *American Economic Journal: Macroeconomics* 2 (4): 1–35.

Diewert, W. Erwin. 1976. "Exact and superlative index numbers." *Journal of Econometrics* 4: 115–45.

Diewert, W. Erwin. 2009. "Similarity Indexes and Criteria for Spatial Linking." In *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, edited by D. S. Prasada Rao, 183–216. Cheltenham, UK: Edward Elgar.

Diewert, W. Erwin. 2013a. "Methods of Aggregation above the Basic Heading Level: Linking the Regions." In *Measuring the Real Size of the World Economy*, edited by Sabra Ledent, 169–96. Washington, DC: World Bank.

Diewert, W. Erwin. 2013b. "Methods of Aggregation above the Basic Heading Level within Regions." In *Measuring the Real Size of the World Economy*, edited by Sabra Ledent, 121–68. Washington, DC: World Bank.

Dykstra, Sarah, Benjamin Dykstra, and Justin Sandefur. 2014. "We just ran twenty-three million queries of the World Bank's website." Center for Global Development Working Paper 362.

Dykstra, Sarah, Charles Kenny, and Justin Sandefur. 2014. "Global absolute poverty fell by almost half on Tuesday." *Center for Global Development*, May 2, http://www.cgdev.org/blog/global-absolute-poverty-fell-almost-half-tuesday.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105 (10): 3150–82.

Feenstra, Robert C., Hong Ma, J. Peter Neary, and D. S. Prasada Rao. 2013. "Who Shrunk China? Puzzles in the Measurement of Real GDP." *Economic Journal* 123 (573): 1100–1129.

Heston, Alan. 2013. "Government Services: Productivity Adjustments." In *Measuring the Real Size of the World Economy*, edited by Sabra Ledent, 413–40. Washington, DC: World Bank.

Inklaar, Robert. 2013. "'Price Levels and Economic Growth: Making Sense of the Revisions to Data on Real Incomes': A Comment." *Review of Income and Wealth* 59 (4): 614–22.

Inklaar, Robert, and D.S. Prasada Rao. 2017. "Cross-Country Income Levels over Time: Did the Developing World Suddenly Become Much Richer?: Dataset." *American Economic Journal: Macroeconomics.* https://doi.org/10.1257/mac.20150155.

Johnson, Simon, William Larson, Chris Papgeorgiou, and Arvind Subramanian. 2013. "Is newer better? Penn World Table revisions and their impact on growth estimates." *Journal of Monetary Economics* 60 (2): 255–74.

Kharas, Homi, and Lawrence Chandy. 2014. "What Do New Price Data Mean for the Goal of Ending Extreme Poverty?" *Brookings Institution*, May 5. http://www.brookings.edu/blogs/up-front/posts/2014/05/05-data-extreme-poverty-chandy-kharas.

McCarthy, Paul. 2013a. "Construction." In *Measuring the Real Size of the World Economy*, edited by Sabra Ledent, 343–68. Washington, DC: World Bank.

McCarthy, Paul. 2013b. "Extrapolating PPPs and Comparing ICP Benchmark Results." In *Measuring the Real Size of the World Economy*, edited by Sabra Ledent, 473–506. Washington, DC: World Bank.

Milanovic, Branko. 2012. "Global inequality recalculated and updated: The effect of new PPP estimates on global inequality and 2005 estimates." *Journal of Economic Inequality* 10 (1): 1–18.

Neary, J. Peter. 2004. "Rationalizing the Penn World Table: True Multilateral Indices for International Comparisons of Real Income." *American Economic Review* 94 (5): 1411–28.

Pritchett, Lant, and Lawrence H. Summers. 2014. "Asiaphoria Meets Regression to the Mean." National Bureau of Economic Research (NBER) Working Paper 20573.

Rao, D. S. Prasada. 2013. "Computation of Basic Heading PPPs for Comparisons within and Between Regions." In *Measuring the Real Size of the World Economy*, edited by Sabra Ledent, 93–120. Washington, DC: World Bank.

Rao, D. S. Prasada, and Gholamreza Hajargasht. 2016. "Stochastic approach to computation of purchasing power parities in the International Comparison Program (ICP)." *Journal of Econometrics* 191 (2): 414–25.

Ravallion, Martin. 2013. "Price Levels and Economic Growth: Making Sense of the Revisions to Data on Real Incomes." *Review of Income and Wealth* 59 (4): 593–613.

Ravallion, Martin. 2014. "An Exploration of the International Comparison Program's New Global Economic Landscape." National Bureau of Economic Research (NBER) Working Paper 20338.

Ravallion, Martin, Shaohua Chen, and Prem Sangraula. 2009. "Dollar a day revisited." *World Bank Economic Review* 23 (2): 163–84.

Rogoff, Kenneth. 1996. "The Purchasing Power Parity Puzzle." *Journal of Economic Literature* 34 (2): 647–68.

Samuelson, Paul A. 1964. "Theoretical Notes on Trade Problems." *Review of Economics and Statistics* 46 (2): 145–54.

Sergeev, Sergey. 2005. *Calculation of the results of the Eurostat GDP comparison with the use of fixity.* Statistik Austria. Guglgasse, Austria, May.

Summers, Robert. 1973. "International Price Comparisons Based on Incomplete Data." *Review of Income and Wealth* 19 (1): 1–16.

United Nations. 1986. *World Comparisons of Purchasing Power and Real Product for 1980*. New York: United Nations and Eurostat.

Van Veelen, Matthijs. 2002. "An Impossibility Theorem Concerning Multilateral International Comparison of Volumes." *Econometrica* 70 (1): 369–75.

Vogel, Frederic A. 2013. "The Ring Comparison: Linking the Regions." In *Measuring the Real Size of the World Economy*, edited by Sabra Ledent, 225–44. Washington, DC: World Bank.

West, Kenneth D. 2006. "Forecast Evaluation." In *Handbook of Economic Forecasting*, Vol. 1, edited by Graham Elliot, Clive W. J. Granger, and Allan Timmermann, 99–134. Amsterdam: North-Holland.

World Bank. 2008. *Global Purchasing Power Parities and Real Expenditures: 2005 International Comparison Program*. Washington, DC: World Bank.

World Bank. 2013. *Measuring the Real Size of the World Economy*. Washington, DC: World Bank.

World Bank. 2014. *Purchasing Power Parities and Real Expenditure of World Economies: Summary of Results and Findings of the 2011 International Comparison Program*. Washington, DC: World Bank.

World Bank. 2015. *Purchasing Power Parities and the Real Size of World: A Comprehensive Report of the 2011 International Comparison Program*. Washington, DC: World Bank.

This article has been cited by:

1. Angus Deaton, Bettina Aten. 2017. Trying to Understand the PPPs in ICP 2011: Why Are the Results So Different?. *American Economic Journal: Macroeconomics* **9**:1, 243-264. [Abstract] [View PDF article] [PDF with links]