Inconsistencies in comparing relative prices over time: patterns and facts

Robert Inklaar, Ryan Marapin, Jop Woltjer and Marcel Timmer

Groningen Growth and Development Centre, University of Groningen

7 April 2021

Abstract

Purchasing power parities (PPPs) aim to measure relative price levels across countries, like inflation aims to measure relative price levels over time. Ideally, the change in PPPs over time should be consistent with relative inflation, but in practice inconsistencies tend to be substantial, which leads to uncertainty about the relative size of economies or about inflation in countries such as China. In this paper, we look for patterns in the PPP data to better understand when and where inconsistency is a more serious problem. We find smaller inconsistencies for more recent PPP comparisons, for countries that are more similar in terms of income levels and expenditure patterns, but larger inconsistencies for consumption products where measurement challenges are larger. We also find that inconsistencies that distort the international income distribution are uncommon. More frequent PPP surveys are unlikely to decrease inconsistency considerably.

Introduction

Statistical programs for measuring purchasing power parities (PPPs), such as the International Comparison Program (ICP), are designed to give an accurate snapshot of comparative price levels across countries at a point in time. We would expect that such snapshots at different points in time bear some relationship to price changes within countries, over time. In theory, small discrepancies due to the use of different weights are to be expected. In practice, though, the change in PPPs over time can be very different from patterns of relative inflation between countries, see e.g., Deaton (2010) or Inklaar and Rao (2017). This points to deeper problems in the comparability of products and prices that are compared in national and international surveys. The discrepancies might have serious consequences for economic research since the choice for a specific snapshot, or PPP benchmark year, can impact estimates of international income inequality or affect the results of entire studies.[1]

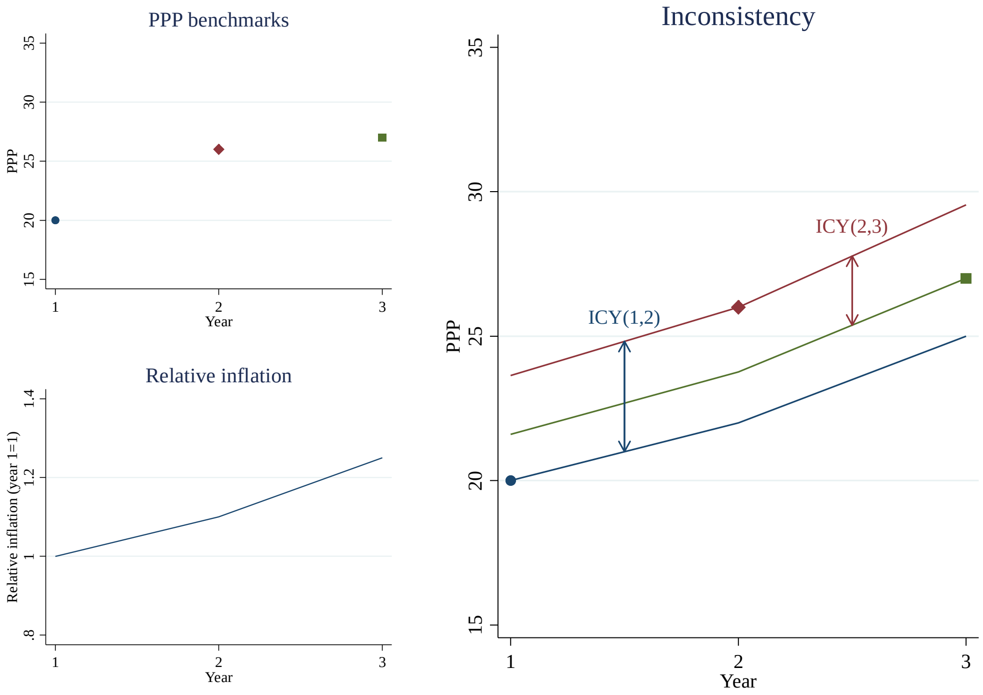
The first substantive discussion of this issue goes back to Krijnse Locker and Faerber (1984). The Ryten report (United Nations Statistical Commission, 1999) also flags ‘incoherence’ (inconsistency in our terminology) as a matter of serious concern for the quality and credibility of the ICP. Summers and Heston (1991) discussed a so-called ‘consistentization’ approach for the Penn World Table (PWT) to reconcile time series of comparative price and income levels compiled based on different PPP benchmarks. More recent examples of methods to reconcile PPP benchmarks and inflation are the approaches of Rao, Rambaldi and Doran (2010) and Hill and Melser (2015). Yet this emphasis on reconciliation skips a step, in our view, because we do not yet have a very good understanding of the scope and features of the underlying problem of inconsistency between PPP changes and relative inflation.

As a step towards remedying this, we document a new set of stylised facts about these inconsistencies in this paper. The aim is to provide a systematic perspective on where and when inconsistencies are larger. This can help inform more formal modelling of the type done by Rao et al. (2010) and provide a view on the uncertainty surrounding PPP estimates in other settings as well, for instance to use these as weights in econometric analysis. By identifying where inconsistencies are largest, we can also help point to specific measurement challenges with the aim of improving statistical practice. We build on earlier research that focused on comparing specific benchmarks, such as Deaton and Aten (2017) and Inklaar and Rao (2017) who compare the ICP 2005 and ICP 2011 results, but we propose a more general framework

[1] See Johnson, Larson, Papageorgiou and Subramanian (2013) and Ciccone and Jarociński (2010).

for analysing inconsistency. This framework is motivated by some of the specific difficulties for comparing prices across countries that Deaton and Heston (2010) discuss.

Figure 1. A stylised example of inconsistency between PPP benchmarks and relative inflation



To frame the issue more clearly, consider the stylised example illustrated in Figure 1. For comparing prices between countries A and B, there are three PPP benchmarks, in years 1, 2 and 3 plotted in the top left panel. We also observe inflation in both countries and can thus plot inflation in country A relative to country B in the bottom left panel. In the right panel, these two sets of information are combined. Starting from each benchmark, an estimate can be made of PPPs in subsequent years using relative inflation. Applying these estimates to each of the three benchmarks leads to the parallel lines shown in the right panel, with the blue line starting from the benchmark in year 1, the red line in year 2 and the green line in year 3.

The distance between two of the lines is our measure of inconsistency, with $ICY(1,2)$ referring to the inconsistency between the PPP benchmark in year 1 and year 2 and $ICY(2,3)$ the inconsistency between benchmarks in year 2 and 3.[2] In the next section, we will provide a more formal framework for measuring inconsistency and comparing the degree of inconsistency in a setting of multiple countries, benchmarks and years. In general, inconsistencies can be due

[2] This general framework is also at the basis of the work of Rao et al. (2010), whose final real income series is based on a weighted average of the three lines shown in the right panel of Figure 1.

to differences in the expenditure shares across countries—with inflation based only on national shares and PPPs on shares for multiple countries—, due to differences in the measurement method for inflation versus PPPs or due to differences in product sampling in national and international price surveys.

The questions we ask of the data are primarily inspired by Deaton and Heston (2010), who highlight the difficulty of comparing prices across disparate countries:

1. *Are more recent global PPP benchmarks more consistent?* Given the more extensive resources devoted to the more recent benchmarks as well as the greater methodological refinement, we would expect increasing consistency over time. At the same time, the more recent benchmarks are more extensive in scope, covering over 170 countries since 2011 versus 115 in 1996 and even fewer before that. This increase in scope could lead to greater difficulties in comparing like-with-like.
2. *Is there more consistency between more similar countries?* Making a global price comparison requires comparing very disparate countries. This will be more difficult if expenditure patterns across products are very different, because a price comparison needs to bridge that gap in spending patterns. It will also be more difficult to compare prices of identical products, as, for example, the quality of housing or of schooling can be very different. This is an important reason why some argue for building a global income comparison by comparing more similar countries in a stage-wise comparison instead a single multilateral comparison (e.g., Hajargasht, Hill, Rao and Shankar, 2018).
3. *Is there more consistency for products that are easier to price and compare across countries?* At a high aggregation level, we expect less inconsistency for household consumption than for other major expenditure categories. At a more detailed level, ICP has always emphasised that certain parts of GDP are “comparison-resistant”, such as public services, housing and construction. There are also a sizeable number of expenditure categories where ICP makes no direct price observations, instead relying on PPPs for other categories or higher aggregates. We expect that it is more likely that the evolution of (e.g.) food PPPs matches relative food prices than that the PPP for public services matches the implicit price deflator for this expenditure category.
4. *Would more frequent benchmark comparisons lead to more consistency?* A longer period between comparisons will lead to larger differences in product samples and expenditure patterns, which could lead to greater inconsistency. Statistical practice is moving towards

more frequent comparisons, with annual PPP estimates across Europe and ICP also shifting from a six-year gap to a three-year gap.[3]

5. *Do inconsistencies distort the international income distribution?* Especially in the papers comparing ICP 2005 to ICP 2011, a major concern was that inconsistencies were larger in lower-income countries, implying larger global cross-country income inequality based on ICP 2005 than on ICP 2011. While any inconsistency makes it hard to establish income rankings of countries, if these inconsistencies vary systematically with income levels, this makes it harder to assess global inequality trends.

We use the most recent version of the Penn World Table, 10.0 (Feenstra et al. 2015) to analyse patterns in inconsistency starting with results from the first ICP comparison for 1970. This version is particularly suited for this purpose because it incorporates the results of the most recent ICP benchmarks, including that of 2017 (World Bank, 2020). If there is a trend over time towards more consistency, being able to compare the ICP 2011 and ICP 2017 benchmarks is important as these have not attracted the type of criticism that the earlier benchmarks have. We supplement PWT data for the 1970–2017 period with information on PPPs and relative inflation for detailed product categories, the basic-heading level of ICP. These more detailed data are used in the construction of the ICP PPPs and cover the period 2011–2017.

We define inconsistency based on Figure 1 as the distance (in log terms) between the parallel lines in the right-hand side panel. Our findings on the questions we formulated above are as follows:

1. More recent ICP benchmarks are less inconsistent, pointing to the importance of improved measurement methods.
2. Price comparisons between countries:
    a. With more similar expenditure patterns are less inconsistent.
    b. With more similar income levels are (frequently) less inconsistent.
3. When comparing inconsistency by household consumption expenditure category, we find that some harder-to-measure product categories such as education and housing have high degrees of inconsistency and some easier-to-measure categories such as food products and clothing are lower. Yet inconsistency is also high for some categories without major measurement challenges, such as furnishing and household equipment.

[3] ICP was due to conduct a global comparison for 2020 before the Covid-19 pandemic hit and that made the requisite data collection much harder. A new comparison is planned for 2021.

4. When comparing PPPs across multiple benchmarks, there is no clear upward trend in inconsistency. We would have expected such an upward trend if differences in spending patterns lead to an accumulation of inconsistencies over time. This result could instead indicate that random factors are predominant in driving inconsistency. However, the modest number of benchmark comparisons make firm conclusions on this hard to draw.
5. The only PPP benchmark where inconsistency varied systematically with income was ICP 1980.[4] As a result, the PPPs from ICP 1980 show a higher degree of income inequality than what is implied by extrapolating PPPs from ICP 1975 forward or from ICP 1985 backwards.

Our paper relates to the work by Rao and Hajargasht (2016) on estimating standard errors for PPPs. The approach taken in that paper (and related literature) is to use the variation in prices for individual items around the (weighted) average price level (i.e., the PPP) as indicative of the uncertainty surrounding the PPP. In our analysis, we try to quantify PPP uncertainty by comparing changes in PPPs to relative inflation. We focus on a different aspect of 'mismeasurement' because inconsistency between PPP changes and relative inflation can be due to mismeasured inflation as well as mismeasured PPPs. Index number problems can also drive inconsistencies—PPPs are estimated using expenditure patterns for multiple countries while inflation is only based on domestic expenditure patterns. At the same time, some of the price variation that Rao and Hajargasht (2016) analyse can be traced to systematic cross-country differences in price patterns. Most notably, the Balassa-Samuelson hypothesis predicts lower relative prices for services than for goods in low-income countries compared to high-income countries.

Our analysis of inconsistency over longer periods of time is especially relevant in a historical context. Since the work of Maddison (2001, 2007), reliance on a single global comparison for 1990 has been the dominant approach (see e.g., Bolt et al. 2018). This is despite growing evidence that this modern price comparison is inconsistent with historical price comparisons (e.g., Woltjer, 2015; Veenstra, 2015) or the price-income relationship that can be seen in every international price comparison (Prados de la Escosura, 2000). Understanding the degree of inconsistency especially over longer periods of time can be helpful to make sense of

[4] Rather than using the official ICP 2005 results, we use PPP data from PWT, which incorporates the adjustments proposed by Inklaar and Rao (2017) to correct for methodological differences and biases.

inconsistencies between modern (i.e., post-1950) price comparisons and historical comparisons, such as by Ward and Devereux (2021).

Conceptual framework

Our goal is to compare the consistency of PPP estimates and relative inflation using price and expenditure information for multiple products and countries. To frame this issue more clearly, let us first consider the price p of an individual product i in a two-country setting, country j relative to country k. The PPP for that product at time t is then defined as:

$$PPP_{ijk}^{t} = \frac{p_{ij}^{t}}{p_{ik}^{t}} \tag{1}$$

Next define the rate of price change over time, π, for that same product between time v and time t in country j (and k):

$$\pi_{ij}^{vt} = \frac{p_{ij}^{t}}{p_{ij}^{v}} \tag{2}$$

Given these definitions, the change in PPP between time v and time t must be equal to the relative rate of inflation between the two countries over the same period:

$$\frac{PPP_{ijk}^{t}}{PPP_{ijk}^{v}} = \frac{p_{ij}^{t}}{p_{ik}^{t}} \Big/ \frac{p_{ij}^{v}}{p_{ik}^{v}} = \frac{\pi_{ij}^{vt}}{\pi_{ik}^{vt}} \tag{3}$$

Again, we are focusing here on the price for a single product, say a bag of rice, so equation (3) must hold.

Comparing the change in PPPs to relative inflation at higher levels of aggregation complicates the equality from equation (3) for three reasons, namely that:

1. Aggregate PPPs and inflation are an aggregate of individual product prices using expenditure weights,
2. PPPs and inflation are sometimes measured in different ways for the same product (category), and
3. PPPs and inflation are based on different samples of products.

Reasons 2 and 3 will be discussed at greater length in the empirical sections but note already here that each of these reasons can help form expectations on where the inconsistencies are expected to be larger. That in turn may provide the grounds for ranking these measurement challenges in order of importance.

The first reason, related to expenditure weights, is explained well in Deaton and Aten (2017, 251), whose exposition we follow here. When calculating inflation across multiple products, the price changes of individual products should be weighted by their share in the expenditure basket in that country. Assuming, for expositional simplicity, that expenditure shares s differ across countries but remain constant over time. Using a Törnqvist index, we can write the difference in overall inflation rate $\pi^{\nu t}$ as:

$$\Delta \log \pi_j^{\nu t} - \Delta \log \pi_k^{\nu t} = \sum_i \left(s_{ij} \Delta \log \pi_{ij}^{\nu t}\right) - \sum_i \left(s_{ik} \Delta \log \pi_{ik}^{\nu t}\right) \tag{4}$$

The Törnqvist PPP at time t can, in turn, be written as:

$$\log PPP_{jk}^{t} = \sum_i \frac{1}{2}\left(s_{ij} + s_{ik}\right) \log \frac{p_{ij}^{t}}{p_{ik}^{t}} \tag{5}$$

Combining equations (4) and (5), we can write the change in PPPs as:

$$\Delta \log PPP_{jk}^{\tau t} = \left(\Delta \log \pi_j^{\tau t} - \Delta \log \pi_k^{\tau t}\right) - \sum_i \frac{1}{2}\left(s_{ik} - s_{ij}\right)\left(\Delta \log \pi_{ij}^{\tau t} + \log \pi_{ik}^{\tau t}\right) \tag{6}$$

The first term in brackets is the log approximation to equation (3)[5], but added to this is the second term, which introduces a systematic difference between the change in PPP over time and relative inflation. This term will be larger when expenditure shares differ more between the countries and when a product has a higher average inflation rate.

As discussed above, the effect of differences in expenditure shares in equation (6) is only one of the three factors that may be relevant in practice. In general, we define the degree of inconsistency between PPP changes and relative inflation, d, as:

$$d_{jk}^{\tau t} \equiv \Delta \log PPP_{jk}^{\tau t} - \left(\Delta \log \pi_j^{\tau t} - \Delta \log \pi_k^{\tau t}\right) \tag{7}$$

We express the inconsistency in logarithmic form so that the measure is symmetric between countries and time periods, i.e., $d_{jk}^{\tau t} = -d_{jk}^{t\tau} = -d_{kj}^{\tau t} = d_{kj}^{t\tau}$. In a two-country setting with two periods, $d_{jk}^{\tau t}$ provides a complete description of inconsistency, but with multiple countries and

[5] The correspondence is only exact in continuous time.

multiple PPP benchmarks, it is useful to define summary measures, as in Inklaar and Rao (2017). Our main summary measure is the root mean squared inconsistency $RMSI$:

$$RMSI^{\tau t} = \left(\sum_j \left(d_{jk}^{\tau t} - \bar{d}_k^{\tau t} \right)^2 \right)^{\frac{1}{2}} \tag{8}$$

Here $\bar{d}_k^{\tau t} = \frac{1}{C} \sum_j d_{jk}^{\tau t}$ is the average inconsistency over the set of countries C. This measure is based on the inconsistencies for a given base country k, but thanks to the symmetry of the inconsistency measure d, the $RMSI^{\tau t}$ measure is base-country independent.

We will also consider the slope coefficient from regressing log income level on inconsistencies:

$$d_{jk}^{\tau t} \equiv \alpha + \beta^{\tau t} \log y_j + \varepsilon_j \tag{9}$$

Here, again, we choose a base country k but the resulting $\beta^{\tau t}$ is base-country independent.

Data

We base our analysis primarily on the Penn World Table (PWT), version 10.0, see Feenstra et al. (2015) for a general discussion of this dataset and www.ggdc.net/pwt for information on this most recent release. Most importantly, PWT incorporates all global ICP PPP comparisons since the first one for 1970 and up to the latest version for 2017. Country coverage has increased substantially over this period, from 16 in 1970 to 175 in 2017, see Table 1 below.[6] The statistical project has also become a much broader exercise, building on a growing body of knowledge regarding both conceptual and practical concerns when comparing prices across countries (World Bank, 2013). In ICP, GDP is built up from the expenditure side of the National Accounts, which means prices are collected for products used for household consumption, for government consumption and for investment. In PWT, estimates for prices of exported and imported products are added to get a complete accounting of GDP.[7] This means that, in addition to measuring the inconsistency between GDP PPPs and changes in the GDP deflator, we can also measure inconsistency at the level of the major expenditure categories.

[6] 176 countries participated, but this includes Bonaire for which complete GDP-level data is not available, see World Bank (2020).

[7] In ICP the exchange rate is used to convert exports and imports to a common currency. PWT relies on the estimates by Feenstra and Romalis (2014) for estimates of quality-adjusted export and import prices.

Table 1 lists all global benchmarks included in PWT, the number of countries participating in each comparison and the number of countries that can be compared across different benchmarks. So, the 'vs. t-1' column shows the number of countries that were in both the current and the previous benchmark (so 16 in both ICP 1970 and ICP 1975), column 'vs. t-2' shows the number that were in the benchmark and the benchmark before that (so 14 in both ICP 1970 and ICP 1980) and so on. It is good to note here that 1996 was not an official global ICP benchmark, but rather a synthetic one constructed for PWT version 6.x based a PPP benchmark for a set of regions in 1993, linked to the OECD/Eurostat benchmark for 1996 (see Heston, Summers and Aten, 2002). And while ICP 2005 is a regular global benchmark, PWT corrects for methodological differences with the subsequent benchmarks and the bias in the linking of the regional comparisons (Deaton and Aten, 2017; Inklaar and Rao, 2017).

Table 1. The number of participating countries in each ICP benchmark and the number of countries when comparing to previous benchmarks

ICP Benchmark	Participating countries	vs. t-1	vs. t-2	vs. t-3	vs. t-4	vs. t-5	vs. t-6	vs. t-7
1970	16							
1975	33	16						
1980	60	27	14					
1985	63	41	26	14				
1996	115	59	51	30	13			
2005	145	99	55	53	32	16		
2011	177	142	110	63	59	32	16	
2017	175	173	140	109	63	58	32	16

Notes: Column 'vs. t-1' lists the number of countries that participated in both that comparison and the previous one, so 16 countries participated in ICP 1970 and those same 16 also participated in ICP 1975. Column 'vs. t-2' compares to two comparisons earlier, so only 14 of the countries that were in ICP 1970 were also in ICP 1980.

Table 1 illustrates that simply comparing the maximum set of countries across benchmarks leads to very unbalanced samples, with more countries covered in more recent years. Especially when trying to establish whether inconsistency has decreased over time, this sample variation can be problematic. But a balanced panel that covers all 7 ICP benchmarks would include no more than 13 countries. To strike a middle ground, we define a balanced sample using the 52 countries that participated in every ICP comparison since 1985.

Table 2 shows summary statistics for both samples. As the table illustrates, the balanced sample becomes less representative over time, as benchmarks after 1985 participation grew in particular among lower-income countries. Of further interest is that there is no clear trend in the average price level (column 2–3 and 6–7) and an increasing trend in the average income level. The main outlier is the ICP 1980 benchmark, which shows higher relative prices and lower income levels than the 1975 or 1985 benchmark. This prefaces one of our findings, namely that the 1980 benchmark was the only one to substantially distort the international income distribution.

Table 2. Summary statistics for every benchmark year for the full sample of countries and a balanced sample.

Benchmark	Full sample				Balanced sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	N	p_Y	p_C	Y	N	p_Y	p_C	Y
1970	16	0.65	0.70	10404				
1975	33	0.79	0.79	10468				
1980	60	0.95	0.98	10065				
1985	63	0.64	0.67	10953	52	0.63	0.68	11555
1996	115	0.73	0.76	13162	52	0.80	0.83	16546
2005	145	0.56	0.60	17182	52	0.71	0.74	22544
2011	177	0.67	0.68	21470	52	0.78	0.80	24836
2017	175	0.59	0.60	21129	52	0.63	0.65	27205

Notes: This table shows descriptive statistics for the full sample and the balanced sample, which includes only countries that participated in every ICP benchmark since 1985. Shown are the number of countries N, the average price level for GDP (the PPP divided by the exchange rate) p_Y, the average price level of consumption p_C and the average GDP per capita level. The price levels are equal to 1 for the United States in every year. The GDP per capita level is in 2017 US dollars ($CGDP_o/POP$ from PWT 10.0).

Most of the questions we ask of the data can be answered at this high level of aggregation. But to establish whether inconsistency is larger for harder-to-measure product categories, we also use more detailed information. Part of the release of the ICP 2017 results (World Bank, 2020) were PPPs for 2011 and 2017 at the so-called basic heading level. Within household consumption, we can distinguish spending on food and non-alcoholic beverages (COICOP 01). Going one step more detailed is spending on food (011), on bread and cereals (0111) and, finally, on the basic heading rice (01111).[8] Matched to this categorisation is information on inflation. Nearly all countries publish consumer price index (CPI) data at the two-digit COICOP level (e.g., food and non-alcoholic beverages) but some even at more detailed levels. In constructing a time series of PPP for the period 2011 to 2017 (see World Bank, 2020, and Inklaar and Rao, 2020) the most-detailed inflation series is allocated to each basic heading. That allows for the analysis of inconsistency at the basic-heading level between 2011 and 2017.

[8] The aim of this statistical definition is to arrive at a fairly homogenous grouping of products. A practical consideration is that it is the lowest level of detail for which information about expenditure can still be compiled.

Results

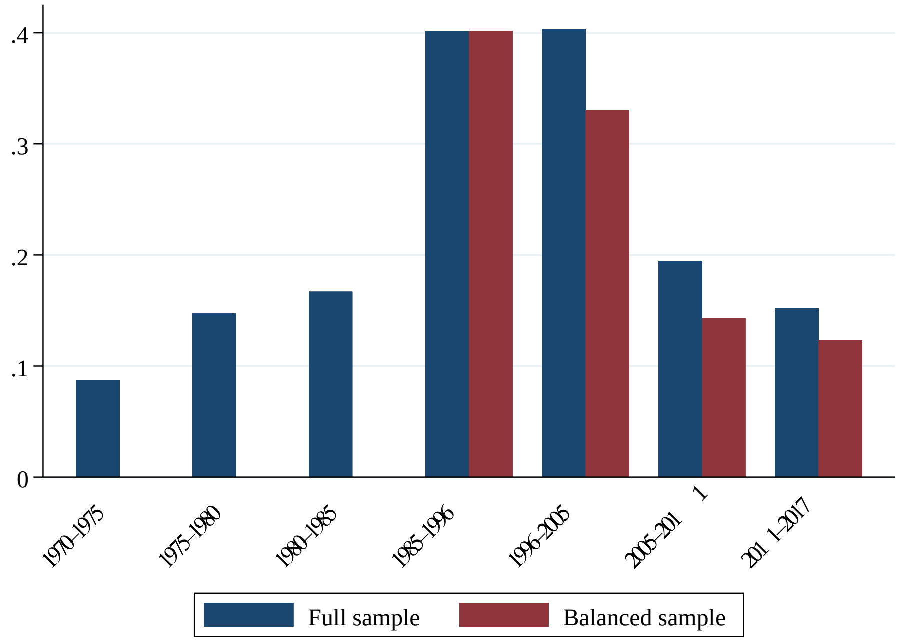
We now turn to answering the questions we set out in the introduction, to assess patterns in the inconsistency data. As a starting point, it is helpful to gauge the overall size and scope of inconsistencies. Across all ICP benchmarks (comparing each benchmark to the next one), we find that the average inconsistency of GDP PPPs, computed as in equation (7), is –0.043. The variation in inconsistency over all countries and years is large with a range between the 25th and 75th percentile of [–0.148,0.071] and a range from the 5th to the 95th percentile of [–0.436,0.320]. This leads to a large average RMSI, based on equation (8), of 0.22. Even without our more detailed analysis that is to come, these descriptive statistics mean that caution is in order, especially when comparing countries that are close in income level as the size of inconsistencies make reversals of income rankings from one ICP benchmark to the next quite possible.

Are more recent global benchmarks more consistent?

Our first question is whether methodological improvements and more extensive resource allocation to statistical programs have decreased inconsistency between more recent ICP benchmarks compared to earlier benchmarks. An alternative possibility is that the rising number and greater diversity of countries covered have made relative price estimations more difficult. Figure 2 shows RMSI estimates for consecutive benchmarks for GDP PPPs, distinguishing RMSI for the full sample and the balanced sample of countries. For the full sample, inconsistencies have increased and then decreased. The increase in inconsistency after the early benchmarks (1970–1985) is remarkable but we should emphasise that country coverage expanded substantially after these early benchmarks. Trying to compare prices across a more disparate group of countries is more difficult, as we discuss below. The decrease in inconsistency since 1985 is similar for the full sample and the balanced sample, which provides support for the hypothesis that statistical improvements have decreased inconsistency.[9]

[9] The largest RMSIs involve the 1996 PPP benchmark (1985–1996 and 1996–2005), which is the only one that is not a proper global price comparison, see Heston et al. (2002). The RMSI for 1985–2005 is 0.29 for the full sample (see Table 4), which is notably lower than the ~0.4 in Figure 2 for 1985–1996 and 1996–2005, but still higher than subsequent comparisons.

Figure 2. RMSI between consecutive ICP benchmarks for GDP PPPs

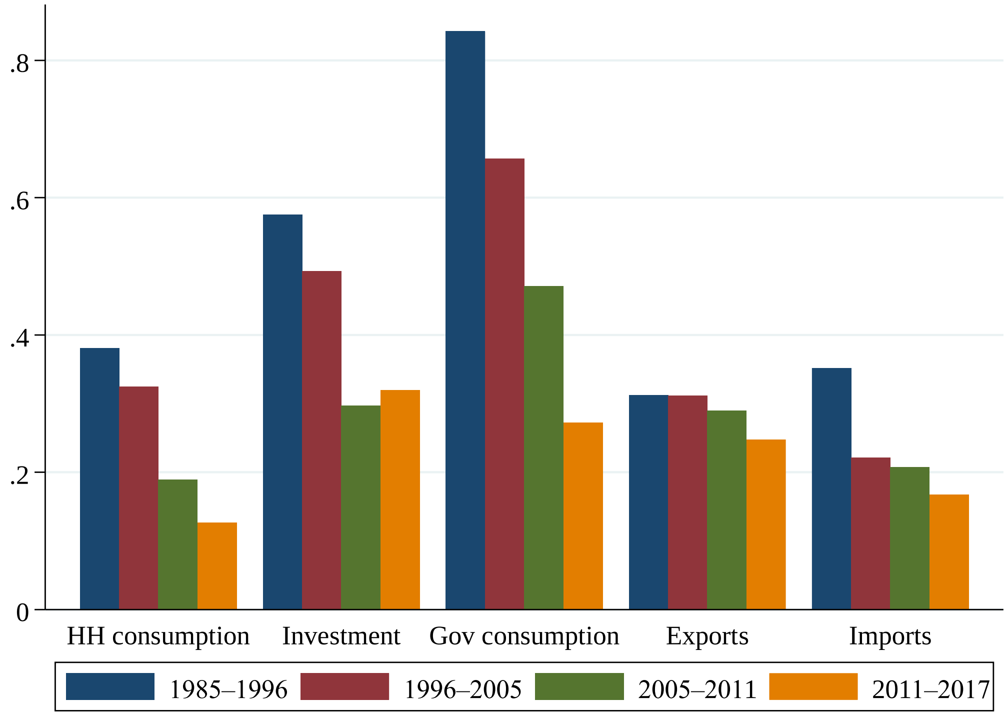


Notes: Each bar shows the root mean squared inconsistency (RMSI), calculated as in equation (8). The inconsistency is computed between two consecutive PPP benchmarks, so the bar labelled '1970–1975' computes the inconsistency (using equation (7)) between the PPPs from ICP 1975 and those from ICP 1970. The full sample includes all countries that participated in the two consecutive PPP benchmarks (see Table 1 for the sample sizes); the balanced sample includes the 52 countries that have participated in every ICP benchmark comparison since 1985.

In Figure 3, the degree of inconsistency over time is shown for each of the five major expenditure categories: household consumption expenditure, gross capital formation, government consumption expenditure, exports and imports. We use data for the full sample of countries and zoom in on the period since ICP 1985 as country coverage becomes broad enough for a meaningful comparison. The PPPs for the domestic expenditure categories (household consumption, investment and government consumption) are based on ICP benchmarks, the PPPs for exports and imports are introduced separately in PWT 10.0 and are estimating using trade unit value data for merchandise trade and following the methodology introduced by Feenstra and Romalis (2014).[10]

[10] The Feenstra-Romalis type export and import PPPs are only available starting in 1984.

Figure 3. RMSI between consecutive ICP benchmarks for PPPs by expenditure category since ICP 1985



Notes: Each bar shows the root mean squared inconsistency (RMSI), see also the notes to Figure 2. The full sample of countries for each comparison is used. 'HH consumption' refers to the National Accounts expenditure category household consumption expenditure, 'Investment' to gross capital formation and 'Gov consumption' to government consumption expenditure. PPPs for these three expenditure categories are based on ICP benchmark data. PPPs for export and imports are from PWT 10.0 based on the method introduced by Feenstra and Romalis (2014) using data for merchandise trade, so excluding trade in services.

The figure shows a very comparable trend for the expenditure categories based on ICP benchmarks, mirroring the GDP-level trend from Figure 2. By the final comparison, between ICP 2011 and ICP 2017, the inconsistency for household consumption is, at 0.13, substantially lower than for investment (0.33) or government consumption (0.28) and even lower than for exports (0.27) or imports (0.17). The decline in inconsistency is more marked for investment and government consumption, though. The decline for the ICP-based categories is monotonic, suggesting a continuous improvement in statistical practice, with particularly strong improvements since ICP 2005. That round marked the start of greater investment of resources by statistical agencies and it had, for the first time, the World Bank in its role as the host organization for coordinating these efforts.

Comparing the downward trend in the RMSI for ICP-based categories with the much less pronounced trend for export and import PPPs is also informative. The data and methods for estimating export and import PPPs has been constant across these years, while ICP methods and data collection efforts have increased substantially. This is a further indication that it is improvements in PPP measurement that led to smaller inconsistency across recent rounds. The higher degree of inconsistency for investment and government consumption compared to household consumption may well be due to the greater prevalence of comparison-resistant expenditure categories, such as construction and collective consumption, a topic we return to in more depth, below.

Is there more consistency between more similar countries?

The second question we ask of the data is for which sets of countries inconsistency is a more prominent feature. The goal of the ICP is to make a global comparison of price and income levels, between the poorest and richest countries in the world, but, as stressed by Deaton and Heston (2010), it is especially when comparing such disparate countries that measurement challenges are greatest. This is for two reasons. First, when comparing two countries, index-number theory argues for the use of expenditure shares of both countries (see equation 5). Yet those expenditure shares are the outcome of consumer decision-making in each country and relative prices will shape spending patterns. Relying on only one country's expenditure shares for estimating relative prices will then impart a substitution bias. Using a Törnqvist index (as in equation 5) or a Fisher index[11] avoids substitution bias by using expenditure share information for both countries, but a consequence is that the comparison is made 'in the middle', i.e., reflecting neither country's spending pattern exactly. However, when expenditure shares are far apart, the 'comparison in the middle' might be a less accurate approximation. This could mean that country pairs with more dissimilar expenditure shares will have greater inconsistency.

A second reason for greater inconsistency for more disparate countries is that the set of products on which the PPP comparisons are based will be more dissimilar from the products included in country consumer price indexes (CPI). Consider two countries where in country 1 fish is the main source of (animal-based) protein and where meat is the more important source in country 2. In the extreme case where country 1 consumes no meat and country 2 consumes no fish, it is not possible to even make a price comparison[12] but for country 1 the CPI will be based on

[11] The Fisher index is the geometric mean of the Laspeyres price index, which compares prices using reference-country expenditure shares, and the Paasche price index, which uses comparison-country expenditure shares.

[12] This extreme outcome is also due to the two-country setup. If country 3 consumes meat and fish, an indirect comparison can be made between country 1 and 2 via country 3.

fish prices and for country 2 is will be based on meat prices. In a more realistic case where both countries consume both sources of protein, some variants of fish may be common in country 1 but not available (or only at relatively high prices) in country 2 and vice versa for some variants of meat. Within a product category, quality differences may lead to similar problems. Housing is a prime example, since a typical house or apartment in a high-income country may be very uncommon in a low-income country and vice versa.

To assess the importance of disparity between countries, we consider two indicators. The first is the (squared) difference between the expenditure shares of countries and we expect that a country pair with a larger difference between expenditure shares, $\delta s_{jk} = \sum_i (s_{ij} - s_{ik})^2$, will exhibit higher inconsistency, d_{jk}. That most closely tests the first reason, where inconsistency arises from index-number challenges. The second is to compute the (squared) difference in income levels and there we expect that a country pair with a larger difference, $\delta y_{jk} = (y_j - y_k)^2$, shows higher inconsistency. While imperfect, this proxy may capture aspects of both reasons.

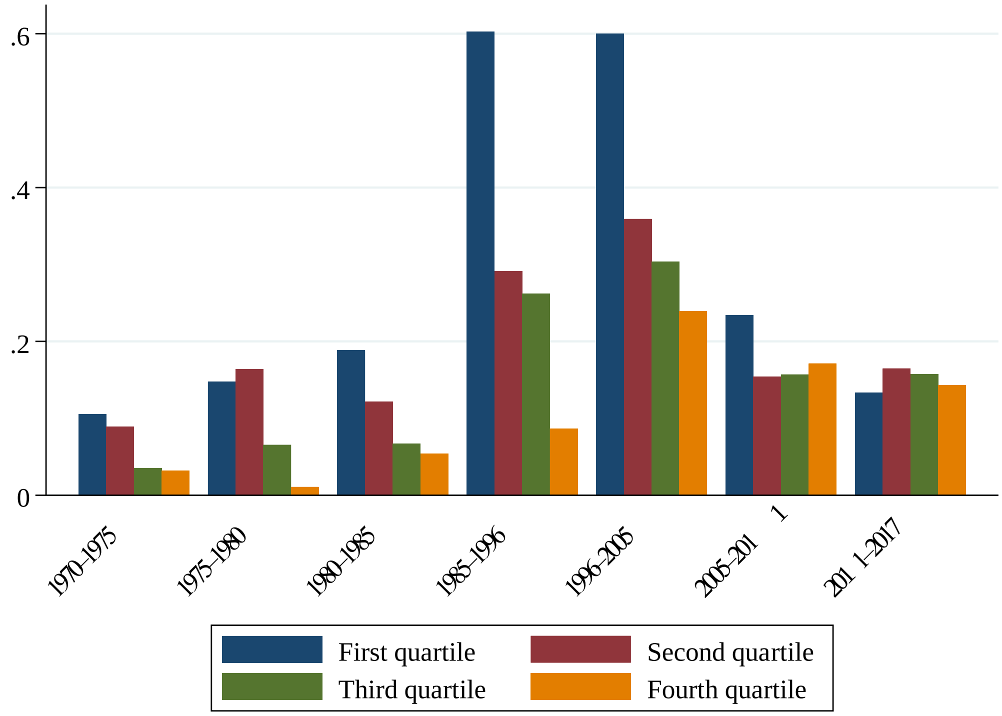
Table 3. Correlation between inconsistency and expenditure share correlations and income level differences for GDP and household consumption

	GDP		Household consumption	
	Expenditure share	Income level	Expenditure share	Income level
1970–1975	0.45*	-0.04	0.44*	-0.06
1975–1980	0.52*	0.21*	0.41*	0.16*
1980–1985	0.27*	0.47*	0.11*	-0.03
1985–1996	0.22*	0.10*	0.20*	0.06*
1996–2005	0.22*	0.04*	0.07*	0.03*
2005–2011	0.21*	0.08*	0.23*	0.03*
2011–2017	0.20*	0.00	0.15*	0.10*

Notes: For each pair of subsequent ICP comparisons (see notes to Figure 2), the inconsistency d_{jk} is computed for all country pairs for GDP PPPs and household consumption PPPs. That inconsistency is correlated with the squared difference between expenditure shares, $\delta s_{jk} = \sum_i (s_{ij} - s_{ik})^2$, and with the squared difference in income level $\delta y_{jk} = (y_j - y_k)^2$. The income level is the (PPP-converted) GDP per capita level, averaged over the two comparisons. Since there are 16 countries that participated in both ICP 1970 and ICP 1975 (see Table 1), the correlations in the first row are based on $\frac{1}{2} \times 16 \times (16 - 1) = 120$ observations. * denotes a correlation coefficient significantly different from zero at the 5-percent level.

Table 3 shows the results of these analyses for GDP PPPs and PPPs for household consumption. We compute the correlation by pair of ICP benchmarks, so, for example, the first row is based on comparing the PPPs from ICP 1970 and ICP 1975. The table shows that inconsistency is systematically higher for country pairs where expenditure patterns differ more, which is in line with prior expectations. Country pairs that differ more in income level also typically show higher inconsistency, but the evidence is less consistent. The correlations are also not large, less than 0.20 for more recent benchmarks. The general pattern of inconsistency thus supports the concerns by Deaton and Heston (2010) that comparisons of more disparate countries are more difficult. Yet the low correlations, especially for more recent benchmarks, indicate that our measures for approximating this disparity are imperfect and/or other (possibly random) factors contribute substantially to inconsistency as well.

Figure 4. RMSI for GDP PPPs by income quartile



Notes: Figure shows the root mean squared inconsistency (RMSI) across consecutive ICP benchmarks, see also notes to Figure 2. For each comparison, the countries are split into quartiles by income level. The income level is computed in each year as GDP per capita relative to the United States and that relative position is averaged over the two years of the PPP benchmarks. These calculations are based on the unbalanced sample of countries.

Figure 4 provides another perspective on the question whether inconsistency is greater when comparing more disparate countries. For each pair of benchmarks, the countries that participated in both are ranked by average income level over the two years. The RMSI is then computed over each quartile of the sample. The figure shows that the RMSI was substantially greater for lower-income countries for the earlier ICP benchmarks, but this pattern is much more muted when comparing ICP 2005 and ICP 2011 and has even disappeared when comparing ICP 2011 and ICP 2017. So, while Table 3 illustrates that inconsistency is a larger concern when income differences are large and/or when expenditure patterns are very different, improvements in price measurement seem to have helped reduce inconsistencies, primarily for lower-income countries. From this analysis, we cannot conclude whether that is due to improved price sampling and measurement for PPPs, for CPIs or a closer alignment of PPP and CPI price samples.

Is there more consistency for products that are easier to price and compare across countries?

As already seen in Figure 3, different expenditure categories show larger inconsistencies than others. Zooming in on more detailed expenditure categories, we know that some are considered "comparison-resistant" such as construction and collective services. In addition, PPPs for some detailed expenditure categories do not rely on direct price observations. Instead, in ICP these are based on PPPs from other categories or higher aggregates; these are referred to as reference PPPs. For example, rather than being directly observed, the PPP for narcotics is based on the PPPs for pharmaceutical products and for tobacco.

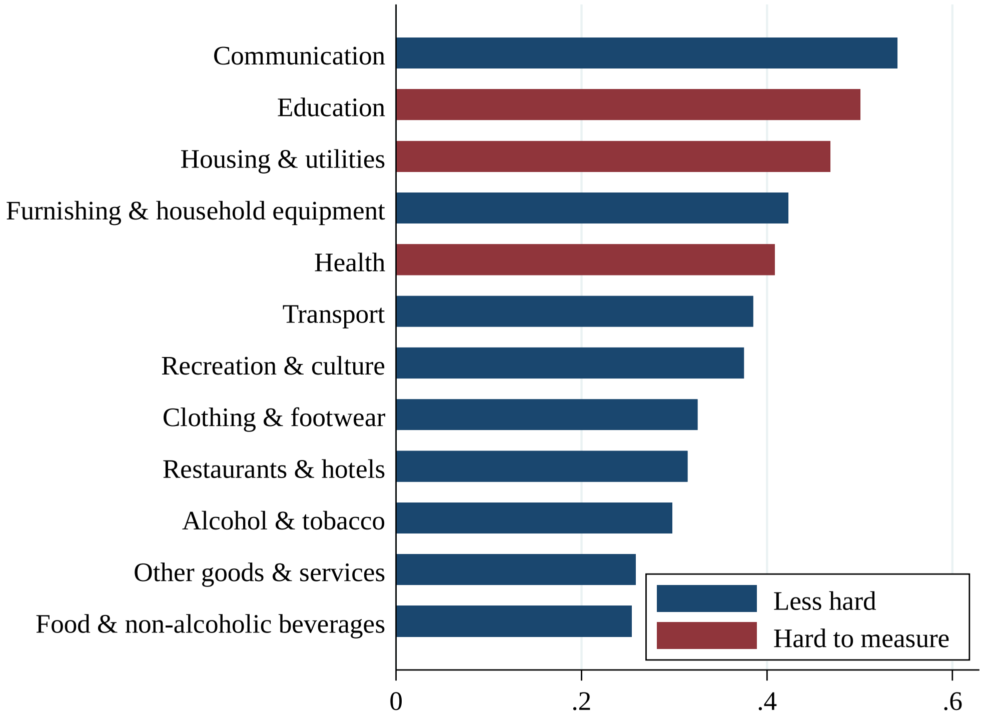
To assess whether there is more consistency for expenditure categories with items that are easier to price and compare across countries, we compute RMSI estimates for the 12 expenditure categories within household consumption expenditure between ICP 2011 and ICP 2017. These are the two rounds of ICP for which we have both PPPs for detailed expenditure categories (basic headings) and inflation series at a more specific level of detail than an overall CPI. Limiting this comparison to household consumption expenditure categories is because these are the categories for which direct PPP measures are most commonly available and because CPI data is typically available at this level. We could also make this comparison for other expenditure categories, but the interpretation is more difficult because there is a larger difference between the approach taken for measuring PPPs and the approach for measuring inflation. For example, ICP relies on the exchange rate for the balance of exports and imports while national statistical agencies would construct export and import price indexes.

In Figure 5 we show the unweighted average of RMSI estimates within each expenditure category.[13] We make a somewhat ad-hoc distinction between categories where price measurement is relatively hard (the red bars) and where it is easier (blue bars). This distinction

[13] The figure is very similar when weighting by expenditure shares within each category.

is based on methodological considerations, as in World Bank (2013), where challenges for measuring the relative price of housing or education are discussed at length. The "hard-to-measure" categories also more frequently rely on reference PPPs rather than direct price measurement.

Figure 5. RMSI by household consumption expenditure category between ICP 2011 and ICP 2017



Notes: Root mean squared inconsistency (RMSI) is computed at the basic-heading level using data from ICP 2011 and ICP 2017. An unweighted average of these basic-heading level RMSIs within each household consumption expenditure category is computed and shown here. Blue bars ("less hard") are those categories for which price measurement challenges are modest; red bars ("hard to measure") cover categories that are sometimes described as 'comparison-resistant.' There, products and price-defining characteristics are hard to define precisely or where no direct pricing is feasible.

The figure suggests some relationship between the RMSI and whether a category is harder or easier to measure. An easier-to-measure category such as food has a low RMSI, while education, housing and health show a higher RMSI. At the same time, the communication category has the highest RMSI, and the furnishing & household equipment category has a higher RMSI than health. A binary easy/hard distinction is not ideal as there are gradations in measurement challenges, even within these broad categories.[14] Yet the current distinction does suggest that inconsistency is a more substantial problem where measurement problems are thornier.[15] We also compared the average RMSI for basic heading categories based on reference PPPs and those based on direct price measurement, which is another perspective on easy versus hard-to-measure. The average RMSI for both groups is very similar with the direct price measurement categories even showing a slightly higher RMSI (0.35) than those based on reference prices (0.32). So, while Figure 5 indicates that more serious PPP measurement problems can lead to more substantial inconsistency (and thus that improving measurement in those areas may help reduce inconsistency), the pattern of inconsistency is more varied than a simple easy-vs-hard to measure distinction.

Would more frequent benchmark comparisons lead to more consistency?

A typical response by statisticians to inconsistencies between PPP benchmarks is to increase the frequency of cross-country price comparisons. Indeed, Eurostat provides annual estimates of PPPs based on a rolling-survey approach where every expenditure category is covered once every two or three years. This ensures that PPPs need not be estimated solely based on inflation information. Increasingly, regional agencies, who coordinate efforts of statistical agencies in different regions of the world, are moving to more frequent cycles. For example, Western Asia's ESCWA has been publishing annual PPPs since the results for 2011. The ICP is also accelerating its benchmark cycle from 6 years (since 2005) to every 3 years—though the Covid-19 pandemic meant the planned ICP round for 2020 was shifted to 2021.

There are broadly two possibilities for what the effect of such more frequent benchmarks could be. First, if there are systematic reasons for inconsistency, such as the term based on differing expenditure shares in equation (6), it could be that inconsistencies accumulate over time and that shortening the time period between benchmarks would lead to a smaller RMSI. Alternatively, if the reasons for inconsistency are largely random and caused by variation in product samples and in inflation rates between products, then it could be that inconsistency remains stable. Such a result would not imply that more frequent benchmark comparisons are useless. More frequent comparisons might be useful to maintain expertise in statistical

[14] For example, PPPs for (imputed) housing rentals are amongst the more challenging to measure, but PPPs for electricity are conceptually much easier.

[15] When considering all basic heading categories within these broader categories as either 'easy' or 'hard' to measure, the RMSI for the 'hard' categories is 0.44 versus 0.33 for the 'easy' categories, a statistically significant difference.

agencies, be relatively cheaper because surveys need not be setup anew and help to track changing product availability more closely.

This question is also relevant for historical price and income comparisons. Maddison (2001, 2007) relied on a single modern benchmark comparison of income levels for the year 1990 and then used data on GDP per capita growth to extend these figures back in time to the 19th century and earlier.[16] This approach was certainly defensible when Maddison first developed his data, but since then there have been increasing efforts to develop contemporaneous income comparison. A good example is the recent work by Ward and Devereux (2021), providing PPP estimates for a set of economies for 1872 and 1910, but see also Bolt et al. (2018) for an overview of such studies. The greater availability of historical income comparisons raises the question how to assess the inconsistency between those historical figures and the original Maddison figures (or other projections from modern price comparisons). If inconsistency tends to accumulate over time, for instance because consumption has dramatically shifted from food to services, the projections of GDP per capita data by Maddison (2001, 2007) over very long time periods is much harder to defend. If random variation dominates, inconsistency could still be substantial, but it would be harder to discount estimates based on modern price comparisons, especially when these are based on higher-quality data than can be used in historical comparisons.

Table 4. RMSI across multiple benchmarks

	Final benchmark:			
	1996	2005	2011	2017
1 benchmark apart (baseline)	0.40	0.33	0.14	0.12
2 benchmarks apart		0.29	0.37	0.17
3 benchmarks apart			0.32	0.35
4 benchmarks apart				0.33

Notes: The table shows the root mean squared inconsistency (RMSI) for GDP PPPs using the balanced dataset for 52 countries. Each row shows the interval between benchmarks, with the first row showing the baseline figures with inconsistency computed based on subsequent benchmarks, so the top-left figure (0.40) is the RMSI when comparing ICP 1985 to ICP 1996. The second row is based on two benchmarks apart, so the first figure (0.29) is based on comparing ICP 1985 to ICP 2005, the second (0.37) based on comparing ICP 1996 to ICP 2011, and so on.

[16] This 1990 benchmark does not correspond to a single ICP PPP benchmark, but is instead based on a variety of sources, including ICP 1985, ICP 1980, Eurostat and OECD comparisons and (via PWT) price comparisons based on expat cost-of-living indexes.

Though it is not possible to assess how moving from a 6-year to a 3-year benchmark cycle would affect inconsistency moving forward, we can look back and assess how a longer time period between benchmarks would have affected inconsistency. This also brings us closer to the time frames for historical income comparisons. Table 4 shows what happens to the RMSI when not comparing consecutive benchmarks. The first row shows the RMSI for the balanced panel of 52 countries since ICP 1985 based on consecutive ICP benchmarks, so our approach so far. The top left figure of 0.40 is the RMSI when comparing ICP 1985 to ICP 1996 and is the same as shown in Figure 2. The second row skips one benchmark, so the first figures in that row (0.29) is the RMSI when comparing ICP 1985 to ICP 2005, the second figure (0.37) when comparing ICP 1996 to ICP 2011, and so on.

Reading this table down the diagonal, so with the same initial ICP benchmark but skipping more benchmarks, does not show a clear trend in the RMSI. For ICP 1985 as a starting point, going to 1996 leads to a larger RMSI (0.40) than skipping to 2005 (0.29), 2011 (0.32) or 2017 (0.33). Starting from ICP 1996 shows a small increase in the RMSI, from 0.33 via 0.37 to 0.35. Another perspective is going by row and then the average RMSI for ‘1 benchmark apart’ is lower than for 2, 3 or 4 benchmarks apart. However, there is no clear difference between 3 or 4 benchmarks apart, so even if inconsistency were to increase with longer time periods between benchmarks, it is not clear whether that trend would continue. So, conversely, whether inconsistency would decrease if ICP benchmarks become more frequent is uncertain. With the relatively small number of ICP benchmarks, though, caution is in order in drawing conclusions. Caution is warranted even more because of the variable number of years between ICP benchmarks, with an 11-year gap between 1985 and 1996 and a 9-year gap between 1996 and 2005 but 6-year gaps since then.

Drawing conclusions relevant for historical income comparisons is also hazardous since the time frames are even longer. This will lead to larger differences in economic structure and spending patterns. Disruptions such as the World Wars may hamper reliability of statistics over time even more than in current times, but the lack of high-quality statistics going back further in time may also raise doubts about the quality of historical income comparisons.

Do inconsistencies distort the international income distribution?

Inconsistency between PPP benchmarks and relative inflation is problematic when trying to assess the relative income level of individual countries, but it is even more worrisome when the entire income distribution changes as a result. This was a main concern when ICP 2011 was released and international income differences were notably smaller than had been expected based on ICP 2005 PPPs that were extrapolated using relative inflation rates (Deaton and Aten, 2017; Inklaar and Rao 2017), i.e., income levels of low-income countries were closer to those of high-income countries than had been expected. In the context of equation (9), this meant that low-income countries had predominantly negative inconsistency estimates since a negative inconsistency on PPPs implies a positive inconsistency on real income levels. High-income countries were on average closer to zero.

Table 5. Regression coefficients of GDP PPP inconsistency on income levels

Benchmarks	Coefficient	s.e.	# of countries
1970–1975	0.039	(0.019)	16
1975–1980	-0.092*	(0.021)	27
1980–1985	0.100*	(0.020)	41
1985–1996	-0.095	(0.054)	59
1996–2005	-0.033	(0.040)	99
2005–2011	-0.025	(0.014)	142
2011–2017	0.014	(0.010)	173

Notes: The table shows coefficient estimates of β from equation (9), so the extent to which countries with higher income levels show greater inconsistency for GDP PPPs. The income level is the level of GDP per capita, averaged between the two benchmarks, relative to the United States. Robust standard errors are in parentheses in the column 's.e'. * denotes a coefficient significantly different from zero at the 5-percent level.

Table 5 shows the regression coefficients on income levels based on inconsistency in subsequent ICP benchmark years. Recall that for ICP 2005, we use the PPPs that are part of PWT 10.0 based on the adjustments proposed by Inklaar and Rao (2017) to address the distortions that the original regional linking procedure had imparted. As the table shows, the coefficient for 2005–2011 is not significantly different from zero, which corresponds to the result of Inklaar and Rao (2017). Indeed, the only coefficients that are significantly different from zero are the two involving ICP 1980. The inconsistency between ICP 1975 and ICP 1980 is negatively related with income level, which implies that international income differences were unexpectedly larger based on the 1980 PPP estimates than based on the 1975 PPP estimates. Going from 1980 to 1985, this pattern was reversed with a positive and significant coefficient of similar size as before. The lack of systematic inconsistencies is comforting and implies that, despite the large inconsistencies for individual countries, the broad cross-country pattern of income differences is typically not distorted.

Conclusions

The topic of this paper is the quality of cross-country price and income comparison benchmarks. In our perspective on this topic, we are close to Rao et al. (2010), who build a statistical model to reconcile inconsistencies between different benchmarks and information about inflation and estimate a ‘consistentised’ real income series. But rather than reconciliation, our aim is to document patterns in inconsistency: has it increased or decreased over time? Is it larger for some comparisons and products than for others? With this exploration, we aim to provide more context to interpreting relative income estimates since the 1970s, point the way to where future measurement efforts could be most fruitfully applied and provide a better understanding of patterns in the data to help underpin more statistical efforts.

One conclusion we draw is that it is likely that improved statistical methods for measuring PPPs have decreased inconsistency. Inconsistency based on ICP PPP data has decreased over time, most markedly since ICP 2005, which saw major investments of resources and methodological improvements. Inconsistency has not changed notably for export and import PPPs, which were estimated based on the same data and methods for the entire period, which is further support for our conclusion. Most of the measurement gains were made in comparing income levels of low-income countries, which has no doubt improved our ability to trace global income inequality and put a firmer basis under the World Bank’s figures for absolute poverty. In a further reassuring result, we show that—with the exception of ICP 1980—inconsistency has not shifted the international income distribution. This means that low-income countries were as likely to see their income level relative to high-income countries improve as deteriorate.

Yet inconsistency remains substantial, with a root mean squared inconsistency of 0.1–0.2. This implies that an adjustment in income levels of 10–20 percent is not uncommon when new PPP data are published. Inconsistency is lower when comparing countries with similar expenditure patterns and at more similar income levels. We also find that, within household consumption, inconsistency is higher in expenditure categories where PPP measurement challenges tend to be more substantial. This suggests that improved measurement methods in those areas could help reduce inconsistency even further.

Finally, we find that increasing the period of time between PPP benchmarks does not lead to larger inconsistency, which points to random variation in product sampling and inflation as a primary factor, rather than a systematic accumulation of inconsistency. This could mean that shortening the period between benchmarks would not lead to lower inconsistency. But while reduced inconsistency may not be an automatic outcome of more frequent international price comparisons, there are still good reasons to support these from a broader price measurement perspective. An important institutional argument is that maintaining the expertise that has been developed over the past 15–20 years in measuring PPPs is easier to maintain and extend when that expertise is more frequently called upon, since procedures remain operational and there will be more overlap between staff trained in these procedures. The PPP programmes run by Eurostat and OECD, that published PPPs at more frequent rates than the ICP, serve as key examples of such sustained expertise.

That institutional perspective is also helpful because there may be domestic spillovers from more frequent international price comparisons. Especially in countries with limited resources, the support from the ICP can help maintain and extend expertise in price measurement, which can also be put to good use in constructing more reliable CPI and other domestic price indexes. As discussed earlier, inconsistency between the change in PPPs and relative inflation need not mean that the PPPs are measured in error, it could be due to domestic inflation measurement problems or deficiencies as well.

A broader conclusion we draw from the analysis in this paper is how hard it still is to make international income comparisons, a conclusion shared with Deaton and Heston (2010). A point in case is the question whether China or the United States has the larger economy. In ICP 2017, the World Bank (2020) data show that the two countries were of approximately the same size, with a difference in GDP level of 0.5 percent. In the same data for 2011, the US economy was 10.7 percent larger than the Chinese economy. Yet given the size of the inconsistencies we discussed in this paper, a cautious person would have said the two economies were approximately the same size in that year as well.

References

Bolt J, Inklaar R, De Jong H, Van Zanden JL (2018) Rebasing ‘Maddison’: new income comparisons and the shape of long-run economic development. Maddison Project Working Paper 10

Ciccone A, Jarociński M (2010) Determinants of Economic Growth: Will Data Tell? American Economic Journal: Macroeconomics 2(4):222-246. doi:10.1257/mac.2.4.222

Deaton A (2010) Price Indexes, Inequality, and the Measurement of World Poverty. American Economic Review 100(1):5-34. doi:10.1257/aer.100.1.5

Deaton A, Aten B (2017) Trying to Understand the PPPs in ICP 2011: Why Are the Results So Different? American Economic Journal: Macroeconomics 9(1):243-264. doi:10.1257/mac.20150153

Deaton A, Heston A (2010) Understanding PPPs and PPP-based National Accounts. American Economic Journal: Macroeconomics 2(4):1-35. doi:10.1257/mac.2.4.1

Feenstra RC, Inklaar R, Timmer MP (2015) The Next Generation of the Penn World Table. American Economic Review 105(10):3150-3182. doi:10.1257/aer.20130954

Feenstra RC, Romalis J (2014) International Prices and Endogenous Quality. The Quarterly Journal of Economics 129(2):477-527. doi:10.1093/qje/qju001

Hajargash R, Hill RJ, Rao DSP, Shankar S (2018) Spatial chaining in international comparisons of prices and real incomes. Graz Economics Papers 2018-03. University of Graz.

Heston A, Summers R, Aten B (2002) Penn World Table Version 6.1. Center for International Comparisons at the University of Pennsylvania (CICUP), October 2002

Hill RJ (2004) Constructing Price Indexes across Space and Time: The Case of the European Union. American Economic Review 94(5):1379-1410. doi:10.1257/0002828043052178

Hill RJ, Melser D (2015) Benchmark averaging and the measurement of changes in international income inequality. Review of World Economics / Weltwirtschaftliches Archiv 151(4):767-801. doi:10.1007/S10290-015-0229-6

Inklaar R, Rao DSP (2017) Cross-Country Income Levels over Time: Did the Developing World Suddenly Become Much Richer? American Economic Journal: Macroeconomics 9(1):265-290. doi:10.1257/mac.20150155

Inklaar R, Rao DSP (2020) ICP PPP Time Series Implementation. 5th Meeting of the International Comparison Program (ICP) Technical Advisory Group (TAG), World Bank, Washington, D.C., 20-21 February 2020

Johnson S, Larson W, Papageorgiou C, Subramanian A (2013) Is newer better? Penn World Table Revisions and their impact on growth estimates. Journal of Monetary Economics 60(2):255-274. doi:10.1016/j.jmoneco.2012.10.022

Krijnse Locker H, Faerber HD (1984) Space and Time Comparisons of Purchasing Power Parities and Real Values. Review of Income and Wealth 30(1):53-83. doi:10.1111/j.1475-4991.1984.tb00477.x

Maddison A (2001) The World Economy: A Millennial Perspective. OECD, Paris

Maddison A (2007) Contours of the World Economy 1-2030 AD: Essays in Macro-Economic History. Oxford University Press, Oxford

Prados de la Escosura L (2000) International Comparisons of Real Product, 1820–1990 Explorations in Economic History 37: 1–41.

Rao DSP, Hajargasht G (2016) Stochastic approach to computation of purchasing power parities in the International Comparison Program (ICP). Journal of Econometrics 191(2):414-425. doi:10.1016/j.jeconom.2015.12.012

Rao DSP, Rambaldi A, Doran H (2010) Extrapolation of Purchasing Power Parities using Multiple Benchmarks and Auxiliary information: A New Approach. Review of Income and Wealth 56(s1):S59-S98. doi:10.1111/j.1475-4991.2010.00386.x

Summers R, Heston A (1991) The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–1988. The Quarterly Journal of Economics 106(2):327-368. doi:10.2307/2937941

Veenstra J (2015) Output growth in German manufacturing, 1907–1936. A reinterpretation of time-series evidence. Explorations in Economic History 57:38-49. doi:10.1016/j.eeh.2015.03.001

Ward M, Devereux J (2021) New Income Comparisons for the late Nineteenth and Early Twentieth Century. Review of Income and Wealth 67(1):222-247. doi:10.1111/roiw.12466

Woltjer PJ (2015) Taking over: a new appraisal of the Anglo-American Productivity gap and the nature of American economic leadership ca. 1910. Scandinavian Economic History Review 63(3): 1–22.

World Bank (2013) Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Program—ICP. World Bank, Washington, D.C. https://openknowledge.worldbank.org/handle/10986/13329.

World Bank (2020) Purchasing Power Parities and the Size of World Economies: Results from the 2017 International Comparison Program. World Bank, Washington, D.C. https://openknowledge.worldbank.org/handle/10986/33623.

United Nations Statistical Commission (1999) Evaluation of the International Comparison Programme. https://pubdocs.worldbank.org/en/164821487203245266/UNSC-30-Session-ryten-report-EN-1999.pdf