

# International Comparison Program

---

## Extrapolating PPPs and comparing ICP benchmark results

*Paul McCarthy*



**6<sup>th</sup> Technical Advisory Group Meeting**  
October 3-4, 2011

Washington DC

## Contents

Background.....	3
The issues.....	3
Extrapolating ICP benchmarks .....	3
Introduction .....	3
Why do extrapolations differ from a subsequent benchmark .....	3
Improving extrapolation methods .....	5
Differences between 2005 and 2011 benchmarks resulting from changes in economic structure.....	5
Differences between benchmarks 2005 and 2011 caused by changes in methodology.....	5
Comparing ICP benchmark results .....	6
Discussion points .....	7
ATTACHMENT A.....	8

## **EXTRAPOLATING PPPs AND COMPARING ICP BENCHMARK RESULTS**

### **Background**

1. Since the 2005 ICP results have been published, various institutions/agencies have extrapolated the results for each of the years up to 2010. They include the World Bank (the World Development Indicators – WDI), the University of Pennsylvania (Penn World Tables – PWT) and the OECD. It is inevitable that the results from the 2011 ICP will differ significantly from those in the WDI and the PWT, particularly given that the global financial crisis in 2008 and 2009 affected different countries to very different extents.
2. The Global Office is also going to face some presentational problems when the 2011 ICP results become available because analysts will want to compare them with those from the 2005 round. There is no simple way of presenting the two sets of results on a comparable basis so the TAG has been asked to advise the Global Office on the best way to proceed.

### **The issues**

3. The two issues are closely related but they need to be resolved separately. History has shown that the “accuracy” of the 2011 ICP results is likely to be judged on the basis of how closely the PPPs match those extrapolated from 2005 to 2011 using one or other of the extrapolation techniques available, with either the WDI or PWT being the likely “benchmark” series. Even in normal economic circumstances, a large gap is likely to arise between the extrapolated series and the 2011 ICP results. The effects of the global financial crisis are almost certainly going to provide an additional shock that will result in the gap being larger than usual and it will have quite different effects on participating countries.
4. The results of the 2005 and 2011 ICP rounds are independent of each other because they are expressed in terms of the price levels prevailing in participating countries in each of these two years. However, that will not prevent analysts trying to compare the results, with a high chance that many will make invalid comparisons that may potentially damage the integrity of the ICP.

### **Extrapolating ICP benchmarks**

#### ***Introduction***

5. Over the past decade or so, various studies have shown that extrapolated series will not match those from successive ICP benchmarks. Conceptually, it can be shown that it is impossible to maintain consistency in real expenditures across both time and space, although some reasonable results may be obtained provided that some very restrictive assumptions are met. In practice, these assumptions are unlikely to apply because they are so restrictive, particularly when countries at different stages of economic development are involved, such as in the ICP.
6. The TAG is asked to consider the issues involved in extrapolating PPPs from a benchmark and provide advice to the Global Office on how to explain the inevitable differences between the 2011 benchmark results and those obtained by extrapolating the 2005 ICP results.

#### ***Why do extrapolations differ from a subsequent benchmark***

7. The 2008 SNA contains a section in the chapter on Price and Volumes Measures (chapter 15) describing a number of the issues involved (see Attachment A). The following list provides a summary of the issues (including some that are not included in the SNA):
  - PPPs can be extrapolated at any level, ranging from the basic heading up to GDP, with the more detailed methods likely to produce better results but with the broader levels more likely to be used in practice because of data (un)availability.

- The products to be priced in the ICP are carefully defined to ensure comparability between countries but the products priced in the time series used in estimating the volumes in a country's national accounts are selected on the basis that they are the most representative products available in a country. In addition, a much broader set of prices is used in a country's time series price indexes than the set that can be included in the ICP.
- The weighting patterns used in a country's time series price indexes are specific to that country whereas those underlying the ICP results are an amalgam of those for the countries participating in the ICP.
- The prices in a country's time series price indexes (e.g. the CPI) are adjusted for quality changes over time and countries do not use common methods to adjust for these changes. For example, hedonic methods are used to a different extent in different countries (or not at all in many countries) with the result that the quality adjusted time series are not consistent across countries.
- Despite the use of SPDs in defining the products to be priced in the ICP, it is likely that the average quality of products priced in lower-income countries is lower than that in higher-income countries.
- An assumption underlying the technique of extrapolating PPPs at the level of GDP is that the structure of each country's economy is similar to that of the numeraire country and is changing in the same type of way over time. In practice, the structures of different countries' economies differ significantly, particularly when developing economies are being compared with a developed economy (e.g. the USA which is commonly used as the benchmark in the extrapolation process).
- Many countries use chain-linked volumes in their time series because of the distortions introduced by using a fixed base year for any length of time. As a result, the GDP deflators for such countries behave differently than those for countries that use the more traditional, fixed-base methods to estimate their GDP volumes.
- Changes in the terms of trade are treated as a volume effect in the ICP but they are included in the GDP deflators (i.e. as a price effect ) used to extrapolate PPPs.
- In the national accounts, very few countries adjust their volumes of non-market services for productivity changes. Therefore, differences in productivity over time in different countries will be reflected in the GDP deflators as part of the price changes, leading to an inconsistency between countries in the deflators used as extrapolators. (Even in the 2005 ICP itself, some regions – Asia, Africa, West Asia – explicitly adjusted their country's PPPs for productivity differences while others did not).
- Countries revise their GDP estimates as firmer data become available, with significant revisions occurring when a country undertakes a "major revision" of their GDP estimates, which generally involves a complete reassessment of the data in the national accounts and the assumptions involved in combining various data sets. As a result, inconsistencies arise between the GDP estimates in a time series compared with those provided for the ICP. For example, comparing the GDP estimates supplied for the 2005 ICP with the latest 2005 GDP estimates available in the UNSD's national accounts database shows that 15 of the 146 countries have revised their 2005 GDP level by more than 10 percent, 19 countries have revised their 2005 GDP by between 5 and 10 percent and 16 have revised it by between 2 and 5 percent. In other words, over one third of the countries participating in the 2005 ICP have revised their 2005 GDP level by more than 2 percent since they provided their national accounts data for the 2005 ICP. Only 19 countries have not revised their 2005 GDP at all. One way of overcoming this problem would be to recompute the real expenditures on GDP applying the 2005 PPPs to the revised national GDP estimates for 2005 so that they are consistent with the GDP estimates provided by countries for the 2011 ICP.

### ***Improving extrapolation methods***

8. It is in the interests of all users of PPPs to have the accurate PPPs for non-benchmark years. The Global Office also has a vested interest in the accuracy of extrapolated PPPs because of the misguided tendency for critics to judge the accuracy of ICP benchmarks based on their relationship with the extrapolated series. Possible means of improving methods of extrapolating PPPs include:

- extrapolating at the most detailed level possible rather than just for GDP; however, experience has shown that a lack of consistent, detailed price data will limit the possibilities;
- adjusting the price extrapolators for any terms of trade effect (e.g. by treating net trade separately from the rest of GDP and using a domestic final demand deflator for this latter component);
- systematically taking account in the extrapolated PPPs of the Balassa-Samuelson effect (which hypothesises that PPPs will tend to rise relative to the market exchange rate in a rapidly developing economy because of strong economic growth being related to higher labour productivity for traded goods compared with that for non-traded goods).

### ***Differences between 2005 and 2011 benchmarks resulting from changes in economic structure***

9. Changes in the expenditure shares (in national currency) for major aggregates in a country indicate that the economic structure of that country has changed significantly between 2005 and 2011. Any large changes need to be verified to ensure that they are indeed due to changes in the country's economic structure and not a measurement error.

10. Deaton and Heston show ratios of Laspeyres to Paasche price indexes resulting from the aggregation process. In all probability, these ratios will change between 2005 and 2011 due to the different products being priced or a different distribution of PPPs or changes in weighting structures (or a combination of all these). The Global Office will need to examine cases where the differences are large, first in data validation, then in explaining the results.

11. The static Penn effect is a widely-observed phenomenon in which the price level index tends to be higher in richer countries than in poorer countries, within any benchmark year. As a result, it is reasonable to suppose that the price level will tend to rise relative to the market exchange rate in a country whose economy is growing rapidly. One reason is that, as a country's relative income level increases, the expenditure share of tradeable goods and services increases relative to that for non-tradeable goods and services. As a result, the country's price level generally increases because the price level of tradeable goods and services tends to be higher than that for the non-tradeable ones (the price levels for the non-tradeables are more closely aligned with a country's wage rates, which are relatively low in a developing country). However, this shift may not necessarily result in an increase in the PLI between benchmarks because the PLI is expressed relative to a base (e.g. another country or a regional average) and the PLI will be affected by any changes in the country's exchange rate relative to that base. However, a useful edit would be to check the countries that change their relative position based on their 2005 and 2011 PLIs and determine whether the changes make sense.

### ***Differences between benchmarks 2005 and 2011 caused by changes in methodology***

12. Extrapolating between benchmarks is also affected by changes in methodology between the two years involved. The major methodological changes in the 2011 ICP compared with the 2005 ICP are:

- Estimates of dwelling rental PPPs will be based on prices and the quantity method in all regions and for linking. In 2005 reference volumes were used in Africa and Asia. However, the PPPs using the reference volume method could be computed for 2011 to for comparison with the quantity method.

- The products priced in the core list will have an impact on regional PPPs. Regional PPPs can be computed with and without core items to determine their impact in 2011.
- Using the important/less important classification will affect the 2011 PPPs. In 2011, the PPPs, etc could be computed without those classifications (as in the 2005 round) to determine the effect of using this classification.
- The global aggregation method proposed in 2011 will produce different results from those that would be obtained from the two-stage method used in 2005. The PPPs based on the method used in 2005 should be computed to determine the effect of this change in methodology.
- In 2005, productivity adjustments were made in three of the six regions (Africa, Asia, West Asia) but the regional linking factors were computed without any productivity adjustments. In 2011, it is likely that some regions will use productivity adjustments but others will not. However, linking factors across all regions will be computed with productivity adjustments included for all regions.
- The construction methodology is changing in 2011 but it is so different from that used in the 2005 round that it will be difficult to compare the effects of the change.

13. Once the final 2011 results have been finalised it will be possible to estimate the effects of most of the methodological changes. However, it is important to emphasise that the differences estimated in this way will provide indications of the effects of these changes rather than precise amounts.

### **Comparing ICP benchmark results**

14. The ICP is designed to compare levels of economic activity across countries, expressed in a common currency, in a particular reference year. The ICP should not be used to compare changes in a country's GDP volume over time, with the times series national accounts volumes for each individual country providing the best data source for this purpose. However, when the 2011 ICP results are released it is inevitable that many analysts will attempt to compare the positions of countries in 2011 with those in 2005, which involves simultaneously studying changes across time and across countries.

15. The only way to maintain consistency simultaneously across time and space is to use a common, fixed price vector to aggregate the basic heading values to GDP. Clearly, the ICP does not follow such a restrictive practice because it would severely compromise the accuracy and usefulness of the ICP results. Directly comparing the ICP estimates of real expenditures for 2011 with those for 2005 is impossible because price levels have not only changed between 2005 and 2011 but they have changed to a different extent across countries. Even comparing the relative positions of countries can be misleading when world (or regional) averages are used as the basis for comparison. For example, in a region such as Asia, whose economic activity is dominated by China, the relationship between the real expenditure for one of the high-income countries, such as Hong Kong, and the regional average will decline between 2005 and 2011, even though Hong Kong's real GDP may have risen appreciably between these two years. The reason is that the regional average real expenditure on GDP has increased even more than Hong Kong's real GDP because of the dominance in the region of the rapidly-growing Chinese economy.

16. The only way of directly comparing results between the 2005 and 2011 ICPs is to have a common reference base for the two benchmark years (e.g. the world average in 2011). If the real expenditures for all countries in each of 2005 and 2011 are expressed in terms of the 2011 world average then direct comparisons are possible. However, it is necessary to make a series of assumptions to enable this link to be made. As is the case with time series volumes, any such comparison is not unique. A different set of relativities between the countries for the two years would be obtained if the results were expressed in terms of the 2005 world average. Using a single country

as the base (e.g. the USA in 2011) is not a solution either because the results would vary depending on the base country (and reference year) chosen.

17. Some simple comparisons are possible but presenting them is likely to lead to demands for other types of comparisons rather than to head off such demands. For example, the relationship between pairs of countries are possible, such as “the real expenditure on GDP for country A was x% higher than that for country B in 2011 compared with y% higher in 2005”. However, comparisons between groups of countries will be more problematical. For example, it will not be sensible to say that the real GDP for country A was x% lower than the world average in 2011 compared with y% lower in 2005 because of the price changes as well as the real changes in the world average between these two years.

### **Discussion points**

- (a) What other issues related to extrapolating PPPs from a benchmark year to later years should be included in any documentation or research studies commissioned by the Global Office?
- (b) The issues associated with extrapolating PPPs and real expenditures from a benchmark year to later years are not well known outside those who use PPPs intensively. How should the Global Office educate the broader statistical community about the problems involved?
- (c) Once the 2011 ICP results are available there will be a lot of interest in comparing them with the results from the 2005 ICP. Some simple comparisons will be possible but many of the more “interesting” comparisons will not be sensible. Should the Global Office discourage all comparisons between the 2005 and 2011 results on the grounds that it does not make sense to compare real expenditures across both time and space or should a list of possible types of comparisons be prepared together with a “health warning” about all others?

## **ATTACHMENT A**

### **PPPs and the national accounts**

15.231 .....The ICP is a very expensive and resource-consuming project and so it provides benchmarks at infrequent intervals. As a result, PPP benchmarks, such as the one from the 2005 ICP, have to be extrapolated using time series from the national accounts of the countries involved. It is interesting to compare the outcomes of an extrapolation with the benchmarks from two sets of PPPs compiled several years apart. In practice, the extrapolated series do not tie in exactly with the benchmarks and there are several reasons for the differences that arise. An important one is the issue of the consistency between the prices used in the time series national accounts and those used in calculating PPPs..... Further, the price and volume structure may change significantly over time in a way not picked up in the extrapolation techniques.

### **Why ICP growth rates differ from national growth rates**

15.232 The method commonly used to extrapolate PPPs from their benchmark year to another year is to use the ratio of the national accounts deflators from each country compared with a numeraire country (generally the United States of America) to move each country's PPPs forward from the benchmark. The PPPs derived are then applied to the relevant national accounts component to obtain volumes expressed in a common currency for the year in question.

15.233 Theoretically, the best means of extrapolating PPPs from a benchmark year would be to use time series of prices at the individual product level from each country in the ICP to extrapolate the prices of the individual products included in the ICP benchmark. In practice, it is not possible to use this type of procedure in extrapolating PPP benchmarks because the detailed price data needed are not available in all the countries. Therefore, an approach based on extrapolating at a macro level (for GDP or for a handful of components of GDP) is generally adopted. Leaving aside the data problems involved in collecting consistent data from all the countries involved, a major conceptual question arises with this process because it can be demonstrated mathematically that it is impossible to maintain consistency across both time and space. In other words, extrapolating PPPs using time series of prices at a broad level such as GDP will not result in a match with the benchmark PPP-based estimates even if all the data are perfectly consistent.

15.234 One of the reasons for differences between GDP time series and PPP benchmark comparisons stems from the definition of a product. As explained in paragraphs 15.66 to 15.67, location is an essential product characteristic in the national accounts whereas the PPP comparisons use average prices of the whole country. Another problem is that the weighting patterns underlying the deflators in the time series national accounts will differ from those in the PPP benchmarks over time. In addition, as noted above, the products priced for the PPPs will differ from those underlying the time series because of the requirements in spatial price indices for representativity within each country and comparability between countries, while in time series the main requirement is for consistency over time. Generally, many more products will be priced for a country's price indices than it is possible to price for calculating PPPs. Finally and often most critically, the prices underlying the deflators in the national accounts are adjusted to remove changes in quality over time and the methods of making such quality adjustments can differ significantly between countries. In particular, the extent of using hedonic methods for adjusting products whose characteristics change rapidly varies significantly from country to country. Electronic products (such as computers) feature prominently in hedonic quality adjustment, although some countries also use hedonics to quality adjust products such as clothing and housing. Comparing price changes in a country that uses hedonics in quality adjusting the price indices underlying its national accounts deflators with those in one that does not do so will lead to potentially large inconsistencies between the benchmarks and the extrapolated series.

15.235 Possibly the single biggest factor that affects the difference between extrapolated GDP series and PPP benchmark results is due to exports and imports. GDP volume measures in the national accounts are unaffected by changes in terms of trade whereas they influence real GDP in spatial comparisons directly. For example, an increase in energy prices results in an increase in nominal GDP. In a spatial comparison, the outcome will be an increase in GDP volumes for energy exporting countries relative to other countries because the net trade PPPs are based on exchange rates, which do not respond to a change in the terms of trade to a significant extent in the short term. The result is that the increase in the terms of trade is treated as a volume effect in the PPP-based benchmark. On the other hand, in the national accounts of energy exporting countries, GDP volumes remain unchanged if the same amount of energy is exported and so the increase in the terms of trade is treated as a price effect, which is observed in the GDP deflator used as the price extrapolator.

### **Non-market services**

15.236 Another area that leads to consistency problems between countries' PPP-based volumes is the group of so-called "comparison-resistant services". They are predominantly (although not exclusively) non-market services, with government services being a major part of the non-market services that have to be priced for PPP projects. The main problems in pricing non-market services relate to the quality of the services being produced and the productivity of the labour used in producing them. One of the conventions used in producing the estimates for the government sector in most countries' national accounts is that the value of output is measured as the sum of the labour and material inputs used in producing the service(s), which involves an assumption that an increase in costs translates into an equivalent increase in output. In addition, an assumption that is commonly made in the national accounts is that the productivity of the labour involved in producing such services does not change over time either. A similar assumption, that productivity is identical in all the countries in a comparison, generally has to be made between countries in calculating PPPs. It is a reasonable assumption when countries at roughly the same level of economic development are involved in the PPP comparison. However, when countries at very different levels of economic development are being compared then the validity of the assumption breaks down.

15.237 The choices faced by the compilers of PPPs are either to assume that productivity levels are identical across countries, even when they are at very different stages of economic development, or to adjust the non-market services estimates in some way to account for productivity differences. Apart from the problems involved in determining an appropriate conceptual approach to adjust for productivity differences between disparate economies, obtaining the data required to make such adjustments also proves problematical particularly when the method involves adjustments based on relative levels of capital intensity in the countries involved. Despite the problems, it is sometimes necessary to make productivity adjustments for non-market services because the problems involved in doing so are rather less than the consequences of assuming equal productivity in all the countries in a comparison.

### **Conclusion**

15.238 PPP-based comparisons of activity levels between countries are an important use of national accounts. Despite the conceptual and empirical difficulties, PPP-based volumes provide a much firmer basis for international comparisons than the commonly used alternative of converting national accounts aggregates to a common currency using exchange rates.