Making Sense of the Results of the 2011 International Comparison Program

Frederic A Vogel*
Consultant, International Statistics
The World Bank
Washington, DC, USA
favogel@msn.com

Abstract

A major outcome of the 2011 International Comparison Program (ICP) is the finding that the world economies are more equal is size when using Purchasing Power Parities (PPPs) because consumption and GDP values in most poor economies are larger relative to those of the United States than previously thought based on ICP 2005 measures that were extrapolated forward to 2011. For example, differences between ICP 2011 benchmark and extrapolated Purchasing Power Parities for 2011 for China and India were 17 and 25 percent respectively. These findings have fueled a debate about the quality of the ICP 2005 benchmark results compared to those from the ICP 2011 and also the quality and relevance of PPP extrapolations in general.

The 2011 ICP included more countries than ever before. In addition, new survey and estimation methods were introduced. Intensive analysis of the ICP 2005 results identified shortcomings in some of the methods and also provided the basis for adapting new methods for ICP 2011. The dilemma is that the continual improvement of methods limits the comparisons of PPPs over time. The purpose of this paper is to provide an overview of the properties of the estimation methods and the possible impact of changes made to ICP 2011 to better understand the relationship between ICP 2005 and 2011 results. Key words: International Comparison Program (ICP), Purchasing Power Parities (PPPs), Gross Domestic Product (GDP)

1. Introduction

The 2011 International Comparison Program (ICP) was the largest ever conducted including 177 countries and economies participating at the full level of the GDP and another 22 for household consumption aggregates (World Bank 2015). This compares to 146 countries providing Purchasing Power Parities (PPPs) at the Gross Domestic Product (GDP) level for ICP 2005. In addition, improved methodology based on an extensive review of ICP 2005 (World Bank 2013) was implemented for the 2011 ICP. While the new methodology significantly improved the results, it hampered comparisons with previous rounds of the ICP.

Because the ICP only provides PPPs about every six years, methods are used to extrapolate PPPs forward from the previous ICP. These extrapolations fueled a debate about the quality of the ICP 2011. Using China, the world’s second largest economy as an example, the 2005 ICP showed its real GDP expenditures to be 43.1 percent of the US. In April 2014, the World Development Indicator (WDI) database provided PPPs for 2011 based on an extrapolation of the 2005 PPPs using national account deflators. The extrapolated PPPs indicated that the China’s economy in 2011 was 72 percent of the US. The ICP 2011 results released in late 2014 showed China’s real expenditures to be nearly 87 percent of the US, a significant change caused by a 17 percent difference in the extrapolated and ICP PPPs. The World Bank (2013) provides an in-depth review of the pitfalls encountered with making such extrapolations, but those numbers were

1 The words country and economy are referred to collectively as countries.
used to challenge the 2011 results. However, the differences between the 2005 and 2011 ICP results are greater than what would be expected when considering the growth rates and general economic situations taking place during that period. This paper provides an overview of the changes in methodology that contributed to the differences between the 2005 and 2011 ICPs.

Given the problems underlying the extrapolation method, this paper mainly provides a review of the differences between the ICP 2005 and 2011 PPPs and the contributing factors. Section B provides the methodology, properties, characteristics, and guiding principles underlying the estimation of PPPs within regions and Section C between regions. Section D provides a review of the statistical variability inherit in the estimation process. The paper concludes with a review of how to understand and make sense of the ICP 2011.

2. Within Region Methodology, Properties, and Guiding Principles

The ICP is not only the world’s largest statistical program, but also the most complex. The World Bank (2013) provides a comprehensive review of the statistical theory and methods underlying the estimation of PPPs. The estimation process involves several stages starting with determining PPPs between countries within regions followed by calibrating the regional PPPs to a global currency. The purpose here is to simply describe the basic methods leading up to the estimation of PPPs and their properties as they affect comparisons between the 2005 and 2011 ICPs. Since its inception, the system of national accounts provides the framework for the ICP. GDP expenditures that are broken down into 155 categories called basic headings of which 110 are for household consumption expenditures. These basic heading expenditures are used as weights in the estimation process.

The most significant aspect of the ICP is recognizing its regional nature and each region’s role in the estimation process to make sense of the ICP results. Countries are grouped into regions mostly based on geography with the exception of the OECD countries. Each region develops its own set of products to be priced within each basic heading. For ICP 2005, 18 countries also priced a separate list of products made up of a composite of products from the regional lists after the regional data collection was completed. These prices were used to link within region PPPs to countries in different regions. For ICP 2011, a set of global core products was provided to all regions, which added relevant core products to their regional lists. This set of global core prices provided by all countries was the basis for connecting PPPs across regions for the 2011 ICP, a very significant change from ICP 2005. The following paragraphs describe the methodology, properties, and guiding principles as they impact the regional and global comparisons.

Methodology. Within each region, the first step is the estimation of basic heading PPPs. This starts with the national annual average product prices and for each product the ratio of its price in each country in its currency to the product price in each other country in their currency. Within each basic heading, the product PPPs are averaged for each pair of countries. The Country Product Dummy (CPD) method was used in ICP regions in ICP 2005 and the Country Product Dummy-Weighted (CPD-W) in ICP 2011. The Eurostat-OECD and CIS regions used the Jevons-GEKS\(^2\) for both ICP periods. These are multilateral methods. (see properties below)

Then, basic heading PPPs within each region are averaged to major aggregates and to the GDP, also based on multilateral methods. For each bilateral pair of countries, basic heading expenditure weights are used to compute a weighted average of the basic heading PPPs for each

\(^{2}\text{Gini-Elteto-Koves-Szulc (GEKS)}\)
level of estimation up to the GDP. First, weights from country A are used (Laspeyres index), then country B weights (Paasche index) to compute two different aggregated PPPs for countries A and B. The geometric mean of these two indexes becomes the Fisher index of bilateral PPPs between every pair of countries in the region. The GEKS, a multilateral method, is then applied to estimate the final within region PPPs, which is the geometric mean of the direct and indirect Fisher PPPs with the base country.

Properties. The resulting within region PPPs have the following properties:

- The PPPs between any two countries at each level of estimation are based on the PPPs between each country and their indirect PPPs with the n-2 other countries in the region—thus are multilateral.
- PPPs are Transitive. The direct multilateral PPP between any two countries equals the multilateral PPPs derived indirectly between the other n-2 countries within the region. The multilateral PPP between India/China is the same as the PPP (India/Lao)/China/Lao).
- PPPs are Base Country Invariant. The PPPs between any two countries are the same whichever country in the region is the base.
- PPPs are free of the Gerschenkron effect. This effect is the bias resulting when high-income countries receive more weight causing an overestimation of the real expenditures of low-income countries. As a result, PPP based real expenditures are not additive. Real expenditures of higher levels of aggregation will not be equal to the sum of the components.

Guiding Principles. The relative price levels and shares of real expenditures between countries from the regional aggregation will remain the same after the PPPs have been calibrated to a global currency. This is referred to as the fixity principle. Another guiding principle was to allow regions to make their own choices of methodology for their regional comparison even though it would complicate the linking methods. For example, three different methods were used across the regions for housing, and government PPPs in some regions were adjusted for productivity. These different methods of choice allowed regions to use what was best for their comparisons even though it complicated the linking across regions (World Bank 2015). As a result, the housing and government PPPs are more comparable within regions than across regions.

3. Calibrating within region PPPs to a Global Currency.

The most significant methodological change for 2011 was how within region PPPs were calibrated to a global currency enabling comparisons between countries across regions. The changes were complex involving several factors. At this stage one has PPPs for every country, but comparisons can only be made between countries in the same region. The next step is to calibrate the regional PPPs to the global level where all countries in the world are linked to the same base currency. This linking of regions has to be done in a way to ensure results are again transitive and base country invariant plus the additional requirement that the within region PPPs between any two countries to their base remain the same after they are calibrated to the global currency—the fixity principle.

Methodology. For the 2005 ICP, 18 countries (between 2 and 4 per region) priced a global list coined the “ring” which was the basis for a separate data collection after the regional collection was completed. These “ring” prices were calibrated to the regional base currency using ring country within region basic heading PPPs. PPPs were computed treating each region as a super
country with the Eurostat-OECD region as the base to obtain between region basic heading PPPs. These between region basic heading PPPs times the within region basic heading PPPs calibrated the regional PPPs to the global currency—U.S. = 1.0. The same method was used in ICP 2011 except that the between region PPPs were based on core products, a subset of which each region included in their respective regional lists and priced by all countries. In other words, all countries were reflected in the between region comparison.

The final stage was to aggregate the global level basic heading PPPs to the different aggregates and the GDP to preserve fixity within regions and maintain the transitivity and base country properties. In ICP 2005, the between region basic heading PPPs, or linking factors, were averaged to each level of aggregation. These aggregated linking factors times the within region PPPs calibrated regional results to the global currency at each level and the GDP. The results were transitive, but not base country invariant. In ICP 2011, the Country Aggregation with Redistribution (CAR) method was used. A direct aggregation or averaging was made across 148 countries following the within region process of first computing the Fisher matrix, then using the GEKs to make the PPPs transitive and base country invariant. In order to maintain fixity, the real expenditures for each country were summed to regional totals, which were then distributed within region according to the regional distribution of relative expenditures.

The aggregation to the GDP was done in stages starting with the 148 countries in Africa, Asia, Eurostat-OECD, Latin America, and Western Asia. Russia participated in both the Eurostat-OECD and Commonwealth of Independent States (CIS) comparisons. Global PPPs for the CIS countries were their PPPs from the CIS comparison (Russia = 1.0) multiplied by Russia’s global PPP; the same method was used in ICP 2005. The 22 Caribbean economies were linked to the Latin American region following a similar process. Recall that the mix of other countries in the comparison impacts the PPPs between any two countries; therefore the goal was to minimize the impact of the additional countries, mostly small islands. Neither the CIS nor Caribbean countries had any effect on the comparisons of the 148 countries.

**Properties** The fundamental characteristic of the process to estimate PPPs at all levels from the basic heading to the GDP is the multilateral nature whereby the PPP directly obtained and all n-2 indirect PPPs determines the final multilateral PPP. This results in the following properties.

- The change in the composition of each region affects the comparability with ICP 2005.
- In ICP 2005 global PPPs at all levels of aggregation were simply the respective aggregated linking factors times the within region PPPs. There was no multilateral estimation across countries in different regions.
- ICP 2011 was a true multilateral comparison with resulting PPPs for each pair of countries bringing in indirect comparisons with every other country in the comparison. However, these PPPs were only used to determine each regions share of the world’s real expenditures, which were distributed back to each region to maintain regional fixity.

**4. Sources of Variability**

The ICP is a statistical construct with several sources of variability as shown in table 2. The purpose is to show the variability inherent in the final PPPs and how they compare to alternative methods. Column A shows the final ICP PPPs for selected countries. Columns (b) and (c) show the variability in 155 basic heading and 20+ published aggregated PPPs. The maximum PPPs are generally for machinery and equipment and the minimum for health and education. The point is the user needs to consider the use to be made of the data and choose the appropriate PPPs for the analysis instead of only using PPPs at the GDP level. Section B contains a brief description of
Instead of selected countries pricing a separate list.

The product lists are region specific. The main change for 2011 is that a global set of core products was developed and made available to the regions to include in their regional lists instead of selected countries pricing a separate list. These core products were also used to receive equal weight in the aggregation to the basic heading PPPs using the CPD method. The dilemma is that products in some countries are widely consumed, but in others may be available but not widely consumed. Also, products widely consumed usually have lower prices. Therefore, the “important” and “available but not important” concepts were introduced for 2011. The CPD-W method used weights of 3.0 for important products and 1.0 for not important. These weights were country and product specific. While a weighting process was attempted in ICP 2005 the results were not consistent and the weights discarded. While difficult to measure the effect, the use of the importance classification would have given more weight to less expensive products leading to increasing the level of final real expenditures.

The product lists are region specific. The main change for 2011 is that a global set of core products was developed and made available to the regions to include in their regional lists instead of selected countries pricing a separate list. These core products were also used to receive equal weight in the aggregation to the basic heading PPPs using the CPD method. The dilemma is that products in some countries are widely consumed, but in others may be available but not widely consumed. Also, products widely consumed usually have lower prices. Therefore, the “important” and “available but not important” concepts were introduced for 2011. The CPD-W method used weights of 3.0 for important products and 1.0 for not important. These weights were country and product specific. While a weighting process was attempted in ICP 2005 the results were not consistent and the weights discarded. While difficult to measure the effect, the use of the importance classification would have given more weight to less expensive products leading to increasing the level of final real expenditures.

The Fisher index which is the direct PPP between countries before applying the multilateral methods. Column (d) shows the Laspeyres/Paasche ratio, which is a measure of the difference in the aggregated PPPs based on which country’s weights were used. As economies become more diverse in their price and economic structure, the ratio increases. As described above, the geometric mean of the two aggregated PPPs becomes the bilateral Fisher index which is not affected by indirect comparisons with other countries. Note that the smaller Fisher PPPs, column (e), compared to the final ICP PPPs imply larger real expenditures for those countries.

<table>
<thead>
<tr>
<th></th>
<th>ICP PPPs CAR</th>
<th>Basic Heading CVs*</th>
<th>Aggregate CVs**</th>
<th>L/P Ratio</th>
<th>Fisher PPPs***</th>
<th>GEKS PPPs, no fixity****</th>
<th>IDBM *****</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>China</td>
<td>3.51</td>
<td>.75</td>
<td>.42</td>
<td>1.17</td>
<td>3.42</td>
<td>3.46</td>
<td>3.50</td>
</tr>
<tr>
<td>India</td>
<td>15.11</td>
<td>1.00</td>
<td>.62</td>
<td>1.39</td>
<td>14.93</td>
<td>14.83</td>
<td>14.97</td>
</tr>
<tr>
<td>Russia</td>
<td>17.35</td>
<td>.49</td>
<td>.48</td>
<td>1.03</td>
<td>16.84</td>
<td>16.50</td>
<td>16.32</td>
</tr>
<tr>
<td>Brazil</td>
<td>1.47</td>
<td>.52</td>
<td>.47</td>
<td>1.29</td>
<td>1.47</td>
<td>1.45</td>
<td>1.49</td>
</tr>
<tr>
<td>S. Africa</td>
<td>4.77</td>
<td>.68</td>
<td>.34</td>
<td>1.19</td>
<td>4.79</td>
<td>4.76</td>
<td>4.93</td>
</tr>
</tbody>
</table>

*Standard deviation of basic heading PPPs/geomean of basic heading PPPs
**Standard deviation of aggregated PPPs/geomean of PPPs
***Direct PPP between the US and selected countries before applying GEKS
****GEKS before applying CAR to preserve within region fixity.
*****Real expenditures are additive

Column (f) provides PPPs resulting if the fixity requirement was removed. Note the smaller PPPs compared to the final ICP numbers imply larger real expenditures without the fixity requirement. Finally, column (g) shows PPPs from the IKLE method, which provides additive results. It is noteworthy that the Gerschenkron impact would be small for these countries.

5. Making Sense of the ICP 2011 Results

There are no expenditures at the product level within basic headings, which means all products receive equal weight in the aggregation to the basic heading PPPs using the CPD method. The dilemma is that products in some countries are widely consumed, but in others may be available but not widely consumed. Also, products widely consumed usually have lower prices. Therefore, the “important” and “available but not important” concepts were introduced for 2011. The CPD-W method used weights of 3.0 for important products and 1.0 for not important. These weights were country and product specific. While a weighting process was attempted in ICP 2005 the results were not consistent and the weights discarded. While difficult to measure the effect, the use of the importance classification would have given more weight to less expensive products leading to increasing the level of final real expenditures.

The product lists are region specific. The main change for 2011 is that a global set of core products was developed and made available to the regions to include in their regional lists instead of selected countries pricing a separate list. These core products were also used to

---

3 Ikle-Dikhanov-Balk Method
calibrate within region PPPs to a global currency. It is difficult to assess the impact of the inclusion of the global core products in the regional comparisons. Since they are global products, they may have higher price levels leading to a reduction in real expenditures. However, the importance classification should have minimized this impact. The overriding impact is that the global core price levels will be more consistent with the within region price levels and provide more robust measures to link PPPs across regions.

The PPPs between any two countries within a region are affected by their respective PPPs with all other countries in the region. Countries within each region are treated equally at each stage of estimating the multilateral PPPs. The PPPs of India and China and the indirect PPP to each country in the region receive equal weight regardless of their economic size in the final GEKS estimation. As a result, a change in the mix of countries in a region can affect the PPPs of all countries in the region. The composition of each region changed for ICP 2011 with each adding and deleting countries. The Latin American region had the most changes with Chile moving to the Eurostat-OECD region, Argentina not participating, and the addition of 9 new countries. These additions included mostly small and poorer countries, which impacted the comparisons between countries like Brazil and Peru for ICP 2011 and with 2005. Comparisons with ICP 2005 should be done with care for countries changing regions.

The estimation of basic heading PPPs using the product price comparisons only occurs within regions. There is no comparison of product prices between countries in different regions. Real expenditures are not additive. For example, real expenditures for construction in Africa and Asia exceed real expenditures for Gross Fixed Capital Formation, which includes construction, machinery, and other products.

Because basic heading PPPs are only computed between countries in the same region, PPPs between countries in the same region are more precise than PPPs between countries in different regions. And comparisons between countries with similar economies are more precise than comparison between dissimilar countries. Standard errors can vary from 5 percent for similar economies to 15 percent or larger for very dissimilar economies.

6. Conclusions

The development of a global set of core products, the use of the importance classification, and the CAR aggregation method greatly improved the quality of ICP 2011 compared to ICP 2005. The use of the core products in both the regional and global comparisons ensured that the within region and between region PPPs were consistent. The choice of methods generally increased the real expenditures. The impact of the CAR method was simulated. In 2005, the impact was to raise real expenditures in Asia by 9 percent and in 2011 by 6 percent. This reduced real expenditures in the Eurostat-OECD by about two percent.

References

Serguev, Sergey ((2014) Unpublished analysis