Some Simple Methods to Validate Basic Heading PPPs

Fred Vogel
Contents

The Analysis ...................................................................................................................... 3
Conclusions....................................................................................................................... 7

Figures

Figure 1. PLI's by country................................................................................................ 8
Figure 2. PLI's by basic heading ..................................................................................... 9
Figure 3. PLI's by country............................................................................................... 10
Figure 4. PLI's by region/country .................................................................................. 11
Figure 5. PLS by country ............................................................................................... 12
Figure 6. PLS by region/country ................................................................................... 13
Some Simple Methods to Validate Basic Heading PPPs

A huge effort goes into validating the prices that enter the estimation of basic heading PPPs. The purpose of this paper is to show that once the prices have been validated, there should be another validation that examines the distribution of the PPPs across the basic headings to find those basic heading PPPs that differ significantly from the rest. This can be done by examining the distribution of PPPs within each country. The review of basic heading PPPs is followed by examining the Paasche-Laspeyres spreads based on the global aggregation of the 129 basic headings to the GDP. The paper concludes with questions for consideration.

The Analysis

The input to the analysis to follow is the matrix of 129 basic heading PPPs for the 146 countries in ICP 2005. The distributions can be viewed graphically using box and whisker plots; a concept introduced by Tukey. Box plots are non-parametric and help indicate the degree of dispersion and skewness in the data, and identify outliers.

Table 1 shows the dispersion of the PLIs across the 129 basic headings for Hong Kong, China, South Africa, and Great Britain. (all PLIs expressed with the US =100) These countries were chosen for discussion because they were also Ring countries which means that their global PPPs were also the scalars used to calibrate the respective within region PPPs to the world level. Price level indices (PLI) are used instead of PPPs so that comparisons can be made across countries. The median PLI for HK is 84.4 which simply means half of its basic heading PLI’s are above that value and the other half are below. Of more interest is the spread between the first and third quartiles and then the minimum and maximum values. For South Africa, the minimum PLI was about 1/7 of the median while the maximum PLI was over 10 times larger than the median. Also of interest are the basic headings with the minimum and maximum values and whether there is a consistent pattern.

---

1 The author is grateful for the calculations and input provided by Min Ji Lee, ICP Team, DECDG, the World Bank.
2 John W. Tukey (1977), Exploratory Data Analysis, Addison-Wesley
Table 1. Variables used to evaluate the dispersion in PPPs by country—Hong Kong, China, South Africa, and Great Britain,

<table>
<thead>
<tr>
<th>Country</th>
<th>BH</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>BH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hong Kong</td>
<td>Compensation of employees</td>
<td>18.7</td>
<td>63.6</td>
<td>84.4</td>
<td>106.3</td>
<td>366.7</td>
<td>Passenger transport by air</td>
</tr>
<tr>
<td>South Africa</td>
<td>Compensation of Employees</td>
<td>11.3</td>
<td>56.7</td>
<td>82.2</td>
<td>99.9</td>
<td>836.6</td>
<td>Passenger transport by air</td>
</tr>
<tr>
<td>Great Britain</td>
<td>Compensation of Employees</td>
<td>42.0</td>
<td>95.9</td>
<td>124.1</td>
<td>138.0</td>
<td>395.7</td>
<td>Fuels and Lubricants</td>
</tr>
</tbody>
</table>

*Source: ICP 2005 computations by Min Ji Lee, ICP Team, DECDG, the World Bank

Figure 1 shows the PLIs by country starting with Iceland which has the largest median value followed by Norway with the next highest median value and ending with Iran which has the smallest value. The box shows the range of PLIs between the 25th and 75th quartile. Note that the median is not always in the center of the box; the distance above or below the midpoint is an indication of the skewness. The distance between the median and first quartile in Great Britain is larger than between the median and 3rd quartile. The dotted line shows the minimum and maximum values of the PLIs for each country.

The length of each box is an indication of the variability around the median—Japan is considerably more variable than the countries preceding it. The ranking of the countries from the largest PLI to the smallest shows that the Eurostat-OECD countries generally have the largest median price levels. However, Angola has the 19th largest median value followed by Equatorial Guinea with the 21st largest. This suggests these two countries be examined in more detail. From the box plot, it is easy to see that both have basic heading PLIs that are about 12 times larger than the median. In both cases, the basic heading with the largest PLI is passenger transport by air.

The maximum values by country show that many countries have values that are 4-5 times larger than the median value-15 have maximum values over 10 times larger than the median. A closer examination shows that the basic heading “passenger transport by air” has the largest value for 98 of the 146 countries followed by “fuels and lubricants” for 29 countries. The remaining maximum values are scattered across 14 different basic headings. This suggests that passenger by air and fuels and lubricants be examined in more detail.
The basic headings with the minimum PLI values for 121 countries were those for government compensation. Nineteen of the remaining 24 were basic headings for medical services. The countries with the smallest minimum PLIs were those from the CIS region and were for the compensation basic headings which were not adjusted for productivity. This points to the need to be more consistent across the regions.

The difference between the minimum and maximum PLIs was reviewed. In 15 countries, the maximum value of a basic heading PLI was over 300 times larger than the minimum value. For one country, Tajikistan, the maximum was nearly 7,000 times larger than the minimum. In every case, the basic heading with the maximum PLI was passenger transport by air, and in 12 of the 15 the minimum value was a compensation basic heading. The other three were medical or dental services.

The distribution of the country level PLIs by basic heading is shown in figure 2 starting with the largest median value of the PLI to the smallest. The basic heading, “passenger transport by air” has the largest PLI followed by “fuels and lubricants.” Both have distributions generally outside the range shown by the other basic headings. This implies there may be a definition or specification problem with these two basic headings, therefore, both should be reviewed in detail. Note that services and compensation have the smallest price levels and the greatest dispersion.

The analysis returns to the review of PLIs by country with a more detailed review of the box plots shown in figure 3. Here, the box is defined as before. However, the dotted lines or whisker range is the 1st and 3rd quartiles minus and plus 1.5X the inner quartile range. The circles show the PLIs in each country that exceed the whisker ranges. Again, many of these outliers come from only a few basic headings which should be subjected to another review.

The next step in this review is to examine the six countries that were the base for the global linking. The linking factors for Africa, Asia, Eurostat-OECD, South America, and Western Asia came from the Ring program, those for the CIS were based on the PPPs for Russia in the Eurostat-OECD comparison. The estimation of the linking factors for each region started by converting the ring prices for each country into a common regional currency. This set of regional prices was used to estimate between region PPPs which were in effect scalars to calibrate regional PPPs to the global level. The 5 sets of linking factors times the regional PPPs provided the global level PPPs. These linking factors for each base country were also their global PPPs (within region PPP = 1.0) and can be reviewed using Figure 4.
Figure 4 is the same as Figure 1 except that the countries are grouped by region. The base countries were Hong Kong, Brazil, Great Britain, Russia, South Africa, and Oman respectively for Asia, South America, Eurostat-Eurostat, Commonwealth of Independent States, Africa, and Western Asia. The largest basic heading PLIs for these countries is either passenger transport or fuels and lubricants. This suggests the outliers appearing in the global PLIs may in part or all be caused by the linking factors. In other words, both the ring and regional data for these two basic headings need to be reviewed in greater detail.

The above analysis showed potential problems with passenger transport by air and fuels and lubricants suggesting that both the ring and regional prices be reviewed again. Without further analysis, it appears that the linking factors contribute to the variability and the resulting outliers

In summary, this simple but powerful tool should be used to evaluate the basic heading PPPs before the global aggregation process begins. While the box plots show there is a great deal of consistency across basic headings and countries, there are outliers that need to be examined.

The next step is to review the Paasche-Laspeyres spreads (PLS) based on the global aggregation of the 129 basic headings to the GDP. Figures 5 and 6 show the PLS as the Paasche index divided by the Laspyres. As before, figure 5 shows the countries ranked by the median of the PLS, figure 6 groups the countries by region. The maximum PLS is 16 for Luxembourg and Tajikistan followed by 12 for Qatar and Tajikistan. A review of the PLS matrix shows that the outlier deviations (8 or above) are caused by three countries, Tajikistan, Krygyx Republic, and Qatar with Tajikistan accounting for the most.

The question needing further review is whether the spreads are caused by different structures of expenditures or price levels. The PLI for compensation for Tajikistan is .088 while it is 167 for Luxembourg; the expenditure shares for individual government consumption are 8.1 and 10.1 respectively. The PLI of .088 is clearly an outlier; it is only about 1/30th of the value of the first quartile. The culprit in this case is the relative price differentials.
Conclusions

This brief analysis shows there are two kinds of outliers—basic headings and countries. The above analysis shows that the basic heading passenger transport by air is an outlier. Fuels and lubricants is also an outlier, but may be the result of the linking factors. The issue is whether these problems are caused by undetected pricing errors, inadequate specifications, or a combination. If the issue cannot be resolved with the basic heading remaining as an outlier, then the issue is what to do about it. When should a basic heading for all countries be filled using reference PPPs when the data are beyond the range of the other basic headings?

**Suggestion for discussion purposes:** Consider using reference PPPs when the outlier exceeds the 1st and 3rd quartiles minus and plus 1.5X the inner quartile range.

The analysis also shows that government compensation is an outlier for the minimum values. However, in this case it becomes an outlier because of the low PLIs in the CIS with one country an extreme outlier. In this case, the issue is what to do with a country’s data for some basic headings? In this example, it appears the data for the outlier countries is also reflected in the aggregation by the PLS.

The outliers shown for PLIs by countries and by basic headings, and the outliers in the PLS are mainly caused by a handful of countries. The issue is how to bring them into the comparison after their data have been reviewed and validated for quality purposes.

**Suggestion for discussion purposes:** Compute the global aggregation without the countries, then repeat the global aggregation with them included and redistribute the results to maintain the fixity of the first aggregation.

Figures 1-6 follow.
Figure 1. PLI's by country

PLI (log scale)

PLI (original scale)
Figure 3. PLI's by country
Figure 4. PLI's by region/country.
Figure 5. PLS by country
Figure 6. PLS by region/country

- Asia
- LAC
- OECD–Eurostat
- CIS
- Africa
- Western Asia