



Effect of new benchmark PPPs on the PPP time series

Bettina Aten

Bureau of Economic Analysis, Washington, DC, USA – bettina.aten@bea.gov

Alan Heston

University of Pennsylvania, Philadelphia, PA, USA – aheston@econ.upenn.edu

Abstract

Since 1968 the United Nations Statistical Office and subsequently the EU, OECD, World Bank, and IMF have periodically carried out benchmark international comparisons of prices and expenditures for countries with widely different economic structures. This research was termed the International Comparison Programme or (ICP). The problem this paper addresses is that when two countries participate in subsequent benchmark countries, their relative GDPs and Purchasing Power Parities (PPPs) will not be the same as would be predicted by movements in their relative constant price GDPs or GDP deflators. This paper examines some of the procedures that have been used since ICP-1970 and ICP-1975 through the latest benchmark for 2011. Because the worldwide ICP estimates have been carried out with gaps of 5 to 10 years with the number of countries taking part beginning at 10 and now over 190, the range of evidence on integrating ICP benchmarks has greatly increased over the period. Both the EU and OECD have removed much of the uncertainty of benchmark comparisons by using more frequent benchmarks. However, the problem remains that world benchmark comparisons are unlikely to be more frequent if carried out in the manner of ICPs 2005 and 2011. This means that other approaches must be formulated to provide interim estimates between benchmarks based on an alternative framework. This framework is in process of being formulated and hopefully this paper will help inform the discussion.

Keywords: Purchasing Power Parities; ICP Benchmarks; Consistency between Benchmarks; PPP extrapolation.

1. Introduction

We discuss a specific economic problem: the consistency of price and real product comparisons over time and across space that is an aspect of a more general problem. The general problem is that when surveys are taken at two points in time they do not produce the same results as when the same phenomena is directly surveyed over time. For example, surveys at a point in time find that self-reported happiness or well-being rises with income, while reported happiness or well-being over time does not necessarily rise with real income over time. Fortunately, we are focusing on measures that are less complex to measure than well-being, though the problems of obtaining consistency in different surveys are no less real.

In the first section we take up a seemingly simple case as illustrated by comparing time to time consumer price indexes (CPIs) and cross section price differences across CPI cities, all using the same price surveys within the United States. In the next section we take up the experience with space and time surveys of the International Comparison Programme (ICP), authorized by the UN Statistical Commission in 1968. The ICP in its 47-year history has undertaken cross country price surveys used to estimate purchasing power parity (PPP) comparisons for consumption, gross domestic product (GDP) and other aggregates. In the final section we ask whether the regional experience of the United States and that of the European Union, the Organization for European Economic Cooperation (OEEC), later the Organization for Economic Cooperation and Development (OECD) and the ICP offer any guide going forward. This is an important question because the ICP in the future must use reduced resources to produce quality estimates in a more timely fashion.

2. Regional Price Differences

The US Bureau of Statistics is responsible for bringing out the CPI on a monthly basis, most prices being obtained by onsite visits to different types of outlets. The price collectors have a checklist of characteristics of prices for items from a given outlet type, for example supermarket chains or convenience stores. They will note the characteristics of the volume seller for an entry level item (ELI) like carbonated soft drinks, checking off the container type and size, whether the brand is national or local brand and so forth. If the volume seller remains the same between months, any price change will be recorded and the price changes will form the basis for aggregating the CPI by detailed groups and more summary categories.

The CPI is reported for 38 index areas comprising metropolitan and smaller urban areas. But the individual price observations can also be adapted to compare price differences across the 38 areas at any point in time, typically bringing all the monthly observations into an annual average. A simple hedonic regression of the log of price on the left hand side against the characteristics of the item will produce coefficients yielding the average price of a given specification and the average spatial price differences for that year. These item price differences can then be weighted to provide estimates of aggregate price differences across consumption for each area in each year.

The same set of prices can thus generate two different time to time indexes: one is the CPI for a city like Chicago based on aggregating individual outlet price changes and the second based on the implicit price change for Chicago arising from the difference of Chicago from the US average in two different years. A major caveat with the latter estimation is that the sample design of the CPI targets consistency over time in its framework, not consistency across areas. This means there are item comparisons with very few observations across some areas, and therefore the spatial price differences tend to have a much larger variance than the CPI.

Table 1 shows the percent differences between the actual price levels for 31 metropolitan areas and 7 urban non-metropolitan areas in the US compared with the extrapolated values using the annual CPI for the same areas.

Table 1. Differences between actual and extrapolated price levels across 38 areas in the US

Percent differences between actual and extrapolated price levels*	2006	2007	2008
Maximum	8.6%	8.6%	5.7%
Minimum	-6.4%	-6.7%	-5.2%
Range	15.0%	15.3%	10.8%
Source: Aten and Reinsdorf (2010, Tables 3a,3b,3c)			
*values are shown using the Geary aggregation and the CPI-Urban extrapolation			

The 5.7% in the column labeled 2008 in Table 1, for example, refers to the maximum difference (across all 38 areas), between actual 2008 estimates and extrapolated 2007 estimates using the CPI, and it refers to the New York suburbs which had a price level of 1.337 in 2008 but an extrapolated value of 1.280. Baltimore has a reverse difference, -5.2%, with 2008 price level of 1.021 and extrapolated 2007 value of 1.072. It should be noted that the CPI for New York combine New York city and New Jersey and Pennsylvania suburbs, and the CPI for Baltimore includes the Washington DC area (Table 3c, Aten and Reinsdorf 2010). However, this is not true in 2007, where the maximum is for San Diego and the minimum for Miami-Fort Lauderdale, yet the differences are still high, 8.6% and -6.7% respectively (Table 3b, Aten and Reinsdorf 2010).

Weighting differences are one source of discrepancy between benchmark interarea comparisons and area-specific CPI changes. The transitivity that is imposed on the spatial comparisons means that the weighting structure is an average of all the areas' weights, whereas the CPI for a particular area will use weights that are specific to that area. In addition, spatial weights change over time along with prices, but CPIs hold the weights constant in order to isolate the effects of price changes. Moreover, the formulas used to calculate spatial indexes are not the same as the formulas used to calculate the CPI.

3. The International Comparison Programme (ICP)

Phase I of the ICP, 1970

The initial round of the ICP aimed to develop a methodology for estimating real product and PPPs across economies with very different economic structures, with a reference year of 1970. The 10 countries agreeing to participate were Colombia, Kenya and India representing developing countries, France, Italy, Japan, West Germany, the UK and US representing the developed economies, and Hungary representing the planned economies. Five of the economies had also participated in 4 binary comparisons with the US for 1950 within the OEEC. These binary comparisons of the US with France, Germany, Italy and the UK for 1950 were extrapolated forward to 1970 and compared with the same binary results for 1970, shown in Table 2 below. Here there is a clear pattern, namely that the extrapolations are all below the benchmark estimates for 1970. Further the differences are larger the faster a country has grown with respect to the US. The observations are few, the caveats many, so while interesting, these results are at best suggestive.

Table 2. Binary comparisons and extrapolations 1950 and 1970

US binary with	Per capita Quantities (US = 100)		
	1950	1970 (extrapolated)	1970 ICP
France	45	68	74
Germany	36	67	74
Italy	23	40	48
United Kingdom	56	57	62

Source: Kravis, Kenessey, Heston and Summers (1975, p.9)

Phase III of the ICP, 1975

By the time of the 1975 Report (Kravis, Heston and Summers, 1980, p.325), there were 16 ICP countries for 1970, 1973, and 1975, and extrapolations from 1970 to 1975 could be made for all but Hungary. Three conclusions were clear. First, the difference between the ICP 1975 estimates extrapolated from 1970 and the 1975 benchmark ICP results was much smaller for the 8 higher income economies than for the 7 lower income countries. Second, most extrapolations were below the actual 1975 results, offhand an unexpected result needing further study. Third, the Philippines and India, which had more experienced statistical offices than the other developing countries at the time, had much smaller differences between the extrapolated and actual estimates for 1975. This could be explained by saying that the national constant price series for these two countries are more reliable than for the other developing countries.

The 1975 ICP also allowed extrapolations price levels from actual 1975 back to 1950 for 8 OEEC countries on a multilateral basis. The largest differences were for Italy and the Netherlands, 14% and 19% respectively, while all the others were 8% or less. All of these extrapolations from 1975 were below the actual OEEC price levels for 1950. Put another way the national growth rates exceeded the implicit growth of the two purchasing power estimates with respect to the US. Because the methodologies were different between the ICP and the OEEC studies this result can be noted but not assigned great importance.

The ICP after 1975

After the 1975 benchmark, the EU and OECD coordinated their own comparisons and helped directly with the African region. The world comparisons were built up from the regions following a common framework, while overall responsibility rested with the UN Statistical Office for the 1980 and 1985 comparisons (60 and 65 countries). The next ICP was carried out between 1993 and 1996 with the final coordination of all participating countries carried out at the World Bank, though never published. Our judgment is that that the 1980 ICP has more in common with 1970, 1973, 1975 and 2005 and 2011, whereas 1985 and 1993 comparisons suffer from the fact that the world results had to be assembled in an ad hoc manner. The 2005 and 2011 ICPs were completed on a much larger scale than previous rounds. This is because the number of countries involved was much higher (146 and 199 respectively), there was institutional coordination by the ICP Secretariat at the World Bank with the regions and countries, there were consistent reviews of the data submitted by the individual statistical offices, and the resources available were much greater than earlier rounds. We will not discuss differences between the 2005 and 2011 rounds because papers by Deaton and Aten (2014) and Inklaar and Rao (2014) have done an excellent job of providing explanations of many of the apparent differences between the two benchmarks.

The OECD experience to 1990

The EU practice since 1990 has been to reduce the residual time-space differences by making annual estimates of PPPs for their member countries. Essentially 1/3 of the prices are collected each year and the other 2/3 are updated to the current year. The OECD is moving to that system, but in 1990 had all surveys every 3 years. Heston and Summers (1993, pp.362-5) sought patterns for the 22 OECD countries between 1970 and 1990. Instead of using the US as a reference country, they only used the price level of the four largest European countries, France, Germany, Italy and the UK, in part because they were in all the benchmarks from 1970. Out of a possible 88 cases if all 22 countries participated in all 5 benchmarks (1990 drops out because it is the reference year) there were 61 ICP observations of extrapolations from 1990 to compare actual benchmarks. In general back-casting of price levels from 1990 shows overestimation of price levels or underestimation of income growth for 45 of the 61 cases. Further in the cases where extrapolations overstate the income growth of the benchmarks all were below 8%, and 13 were under 5%. By contrast 11 of the understatements of income growth were greater than 10%, 13 were 5-10%, and less than half were under 5%.

It is again worth noting that the US is just another country in these calculations. One exercise was to examine the pattern of price levels of countries over the 5 ICP benchmarks. The country with the most variation over the 20 year period was the United States, beginning in 1970 at 140.6 with respect to the European 4, dropping to 104.6 in 1975 and 86.4 in 1980, a period accompanying the freeing of exchange rates around the world. In 1985 the US price level was back up to 144.3 and down again to 82.9 in 1990. When the US is the single reference country it makes it appear that there is much variation in the other countries and of course none in the US. The probable reason that the US price level shows such variation is the asymmetrical role that the dollar plays in the international monetary framework.

4. Conclusion

It is clear from the EU and OECD experience that the more frequent the comparisons across countries, the less difference between extrapolations and benchmark. However, even annual comparisons using the same prices, as in our example of the US CPI, may reduce the differences, not eliminate them. ICP experience has also shown that there are gains to disaggregating extrapolations. For example, breaking down GDP into Domestic Absorption and the Net Foreign Balance is important for countries with one or two dominant exports or imports.

Looking forward, the United Nations, International Monetary Fund, World Bank, OECD and other country groups believe the ICP needs to develop methods to produce more up to date results on a more



frequent basis. The early ICP experience showed a tendency for simple extrapolations based on national growth rates to understate the relative growth of countries implicit in benchmark ICP results. However, the understatement has lessened as more countries have been included. In short, there seems no obvious model for updating benchmark ICPs on a long term basis. The OECD has undertaken to incorporate some non-members within their framework of continuing surveys. Non-OECD countries have found that periodic benchmarks have been dysfunctional because of the overall work load and timing of staff resources. For these countries there is a need for a framework at the global level to keep the statistical offices involved with providing ICP information to regional and global coordinators on a continuous basis. This is the work in progress on which the ICP Secretariat is now involved.

References

Aten, Bettina and Marshall Reinsdorf (2010). Comparing the Consistency of Price Parities for Regions in the US in an Economic Approach Framework. Bureau of Economic Analysis Papers, August. (<http://www.bea.gov/research/topics/regional.htm>)

Deaton, Angus and Bettina Aten (2014). Trying to Understand the PPPs in ICP2011: Why are the Results so Different? NBER Working Papers, no. 20244.

Inklaar, Robert and D.S. Prasada Rao (2014). Cross-country income levels over time: did the developing world suddenly become much richer? GGDC Research Memorandum 15.

McCarthy, Paul (2013). Extrapolating PPPs and Comparing ICP Benchmark Results. World Bank (ed.), *Measuring the Real Size of the World Economy*, Chapter 18, World Bank, Washington DC.