The 2005 ICP Benchmark, PWT and Some Health Warnings

Angus Deaton and Alan Heston[1]

Introduction:

This paper is part of a larger effort to place the International Comparison Program (ICP) in the context of macro-economic modeling, and to provide the sort of information that can act as a bridge between producers and consumers of data. The emphasis in the paper is the 2005 round Report of the ICP that was made final in May 2008, while the larger paper will contain an extensive discussion of the theoretical rationale of the spatial and temporal comparisons across countries. The 2005 ICP involves 146 countries, and collection of at least a million prices for specific goods and services making it perhaps the largest single coordinated international statistical activity ever undertaken. It is a rich body of data and the results should be of interest to a wide group of researchers. The first part of this paper describes some features of the 2005 ICP that should be understood in comparing its picture of the world economy with the picture of that same world economy in 2005 as drawn in the World Bank's Development Indicators (WDI), the IMF's World Economic Outlook (WEO), or the Penn World Table (PWT). For example, the earlier world-view had a smaller spread between the poor and rich countries than does the 2005 ICP, from which a hasty inference would be that inequality between countries is higher than we have been used to thinking. That conclusion may or may not be true, but because the methodology of the 2005 ICP has some new features explained below, one would have to apply the old methods to the new data to really make judgments about the range of income spreads across countries.

The second part of the paper focuses on the way that benchmark ICP estimates have been extended over time and space to non-benchmark countries, as in PWT or the WDI. The growth rates from these data sets have been widely (some would say, excessively) used to examine long-term economic development as well as turning points in economic activity. Some uses have been more thoughtful than others and recent replication studies suggest it is *very important* that users understand that there are serious

[1] Princeton University and the University of Pennsylvania. This work is partly supported by NSF grant SES-0648769-001.

health warnings associated with the numbers. This paper focuses mainly on PWT extensions of the ICP benchmark estimates but the PWT caveats also extend to other series.

Before turning to Part I a summary of the new view of the world economy used by the Bank and IMF is provided in Table 1. Column 1 provides the results of ICP 2005 and Column 2 the totals in the WDI of the World Bank, aggregates that are similar to PWT though there would be differences for individual countries. Both Columns 1 and 2 are much closer to each other than to exchange rate totals as given in Column 3, a result found in all rounds of the ICP since 1970. What has caught the attention of many commentators is the smaller share of a smaller World GDP of Africa and particularly Asia, the latter accounted for mainly by China and India. (Note also that Japan, Korea and Mexico are in the OECD in Table 1). We turn now to the implementation of the 2005 ICP and discuss how the results were obtained and why they could differ from earlier estimates.

Table 1: The size of the world economy

GDP 2005 ($ billions)	ICP 2005 GDP @PPP	Previous GDP @PPP WDI	GDP @ ExRates
World-146 countries	54,975	59,712	44,306
High income: OECD	31,422	31,726	33,342
Africa	964	1,264	486
Asia and Pacific	10,971	16,367	4,221
CIS	2,269	2,171	970
South America	2,698	2,911	1,411
West Asia	1,158	932	588

Source: Taken from Table 1 of the ICP 2005 Report

I. Some Conceptual and Practical Issues in Implementing ICP

The basic framework of the ICP has remained the same for the past 40 years, namely detailed price comparisons for specified items on the expenditure side of the accounts. These are put together into purchasing power parities (PPPs) for 155 basic headings. The national accounts on the expenditure side are similarly split into 155 headings providing the inputs into aggregations like food, consumption, and GDP. The devil, of course, is in the details. Fortunately the 2005 Round has a number of new or improved features

including an online Handbook that covers many of the points of this section in more detail than possible here, and readers are encouraged to explore this source written by a variety of experts in the field. In this section we cover 3 broad topics, market and non-market price comparisons, linking the regions, and aggregation issues.

One important practical constraint of the comparisons is the politically convenient practice of ***fixity***, not exactly a four letter word, but a real roadblock for international organizations in constructing world comparisons. Fixity began with the EU and was the requirement that no relationship at the aggregate or sub-aggregate level obtained in comparisons between the member countries be changed when their expenditure and price data were merged with other countries. When the EU and OECD made joint comparisons for 1980, fixity was maintained within the EU, and the joint EU and OECD comparisons. In the world comparison for 1980 other countries were treated individually but by 1985, all regions expressed a desire for fixity of their regional comparisons. Similarly country groups joined to the EU-OECD comparisons, such as associate EU and CIS countries, have their positions in the world fixed by their relation to the EU-OECD. As will be discussed, this posed several constraints on the design of the 2005 ICP comparison, including development of methods to link the regions.

Four issues are taken up in this section beginning with the concept of national average prices for market priced goods and services. A recommendation is made that the matching approach using national average prices be modified to accommodate a more hedonic approach to spatial price comparisons. The second issue is the treatment of the services of housing. The third issue is how you compare non- priced services where there is no market, important examples being publicly provided education and collective government services. The fourth issue is given the fixity constraint, how do you link the results in each region to obtain a global comparison, a major objective of international organizations?

A. National Average Prices

The 2005 ICP followed the practice of earlier rounds in using annual average national prices for a specification as the elemental building block for regional and world comparisons. For a particular specification like white eggs, this has meant obtaining observations in at least 8 outlets in one or more time periods and urban centers. This is

essentially a matching in contrast to a hedonic approach. Adjustments are then made for date of collection and a factor to adjust for rural or non-sampled urban areas to obtain an annual average national price. Rural prices are collected in many African countries and some countries like India have regular rural price surveys from which some price information can be extracted for the ICP. But there remains a strong urban orientation of ICP prices. It seems time to rethink this mantra of national average prices, and consider alternative ways to make such price comparisons. Three reasons for this are briefly described.

Representivity

The EU developed the concept of representivity as a characteristic that countries should assign to the items they price. An item was designated as representative, available, or not representative. The purpose of this devise was to avoid comparing prices of items uncommon in the basket of one country with an item common in another, the logic being that the price of the uncommon item would be high and bias the comparison. For any basic heading, a binary comparison is made between each pair of countries, where representative items in both countries would receive full weight in the comparison, and items representative in one country but only available in the other would receive less weight. Other combinations including items representative in one country but not in the other are excluded from the comparison. The result was that some price information is not used the comparisons; and a parity might not be estimated for a particular heading even though prices were available in both countries for the same item. To obtain parities between all possible binary comparisons between the member countries, the EKS method was used to insure transitivity.[2]

In the 2005 benchmark it was planned to use representivity of items in regions other than the EU/OECD with the goal of not comparing uncommon expensive items in one country with more available lower-price items in another. But, apparently something

[2] A full discussion of this procedure is contained in the Handbook where an alternative to the EU procedure is recommended, the CPRD method. This builds on the country product dummy (CPD) method of Robert Summers (1973), a simple hedonic regression that in recent applications has been expanded to take account of other characteristics of items, like brand or size, and characteristics of the market, like outlet type or urban/rural. The CPRD method systematically incorporates information on representivity and in practice performs better than EKS in estimating basic heading parities. For details, see Ch. 11 in the Handbook under Resources on Bank site: http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/ICPEXT/

was lost in translation of the concept because it proved non-operational in the other regions with the possible exception of South America. In comparison with the ICP rounds from 1970 to 1993, there were many more resources available to allow review of the distribution of individual price quotes within countries, and to evaluate average prices across countries at regional meetings. As a consequence it was possible to eliminate some prices that did not meet specifications or to eliminate items that appeared difficult to compare for most countries in a region. A conclusion from this experience is that a method that has been found acceptable in the EU/OECD has to be modified if it is to be applied in other parts of the world.

The Quality Question

A representative item can be thought of as a widely available item with a substantial volume of sales. However, there are many items that cannot be specified narrowly all over the world, like a women's blouse, which leaves scope for some countries pricing lower quality items than others. Certainly the presumption in earlier rounds of the ICP was that low quality items were often priced in less versus more affluent countries, with a tendency to overstate the output in poorer countries. One reason this might naturally occur is that the average quality of goods entering into the CPI tends to be lower in poorer countries and their outlet sample is likely to be closer to 'Dollar Stores' in the US than to the median U.S. outlet. If the result is that lower quality items in poorer countries were matched with higher quality items in richer countries, it would make their price levels too low. In the past this has been offered as a reason the ICP may have overstated GDP levels in poorer countries.

However, in the 2005 Round there are offsetting factors that have been much stronger than in previous ICPs. Many of the qualities available and in the CPIs of poorer countries are not available in higher income countries, while the qualities in the CPIs of richer countries can also be found in poorer countries. Also, the higher quality items are frequently international brands while regional or brand-less products are more important for lower quality items. The consequence is that higher quality items tend to dominate the actual list of items compared in the ICP, which in 2005 was made more likely because the initial specifications were drawn heavily from the EU-OECD. Many of these items will not be in the CPIs of poorer countries nor necessarily available in the outlets

normally sampled in their CPIs. And this effect may be amplified because CPI outlet and item samples in many developing countries are decades out of date. The consequence is that prices were often collected in higher-end outlets with the effect of raising price levels of poorer countries. The reason this was more likely in 2005 is that there was a much closer review of prices across countries, so that frequently international brands were priced in say, China, because they were available, even if mainly in high-end outlets. To the extent this happened, it would have the effect of raising parities in poorer countries, making them appear to have less output than in fact they do, and to increase the measured gap between rich and poor.

These two quality effects are in opposite directions and the net impact on comparisons is likely to raise price levels more for countries that do not have much local production, e.g., a small country in Africa, where low import volumes, high transport costs, and tariff and non-tariff barriers, may produce high prices. By contrast a large country in Asia is likely to have many domestically produced alternatives, leading to fewer and more competitively priced imports. However, even in large countries like Brazil, China and India that have great variety in outlets and wide range of prices, choice of outlets can have a significant influence on price.[3] More research is needed to evaluate what is the net effect of the quality factor on the overall comparisons. What does seem clear is that compared to earlier rounds, more international brands were priced by the 2005 ICP in higher-end outlets than in earlier rounds. Other things equal, this would lower the relative income of poorer countries in 2005 compared to the pricing practice in earlier rounds.

The Small Country-Big Country Problem

When preliminary results were released of the 2005 round, one point that concerned many observers was that China and India were lower by 40% per capita than in the previous WDI estimates, and also PWT. The estimates are provided in Table 2

[3] In China this has been the case. China agreed to participate in the ICP in the 1993 comparison but on a limited level, namely providing only urban prices. In 1993 the plan was to compare Shanghai with Tokyo and Guangdong with Hong Kong; the Shanghai comparison was never made public but the Guangdong-Hong Kong exercise was completed and described in the publication of ESCAP(1999). For the apparel and footwear grouping, the price level in Guangdong was 64% higher than in Hong Kong in 1993, an improbable result unless it was due to pricing international brand names in upper-end outlets.

where per capita estimates are provided in columns 1-4 from ICP 2005, from the WDI, PWT, and on an exchange rate basis. The last two columns simply express columns 1 and 2 as total GDP for 2005.

One early argument was that price collection in China was organized in a way that provided prices that were too high because of some combination of choice of brand, outlet or center for collection. The price collection by China in 2005 took place in 11 cities and their immediate surrounding areas that had some rural and some urban characteristics. In their review of the Chinese results for the purpose of producing new poverty lines Chen and Ravallion (2008a, Figure 1) concluded that the provinces of 11 ICP cities represented the range of urban poverty lines in all provinces, and encouraging finding. which is done on the basis of rural and urban poverty price levels that they independently estimated by province for 2002. Their conclusion is based on rural and urban poverty price levels that they independently estimated by province for 2002. In that study the 11 urban provinces used in the ICP had an average poverty line of 1243 yuan, which may be compared to an all province urban average of 1195 yuan, and an average rural line of 849 yuan. This line was based upon actual region specific food bundles and so it can be interpreted as price differences between rural and urban areas for a poverty bundle. It is clear that the 11 ICP provinces were only slightly more affluent than all cities.

Table 2: Comparison of Results of 2005 ICP with prior Estimates

	GDP per capita,		PPP	GDP pc US$	GDP	PPP (bln)
	ICP '05	WDI '05	PWT 6.3	ICP '05 EXR	ICP '05	WDI '05
	1	2	3	4	5	6
China	4,091	6,760	6,637	1,721	5333.2	8818.6
India	2,126	3,452	3,536	707	2341.0	3779.0
Japan	30,290	30,736	27,726	35,604	3870.3	3927.3
United States	41,674	41,890	41,674	41,674	12376.1	12416.5

However, it is unlikely that the 11 cities represented the price levels of urban centers in their provinces. For example, Chen and Ravallion report that National Bureau of Statistics (NBS) chose those cities because they were most likely to have outlets carrying the types of products and brands in the ICP specifications. Chen and Ravallion

also note that the rural areas were closer to what would be suburbs than rural areas, and of the 1700 outlets sampled (an impressive number in ICP practice) about 22% were in these 'rural' areas. They conclude that for their purpose a downward adjustment of 35% is required in the 2005 results for China for purposes of approximating prices in rural areas in their poverty analysis.

Price differences between the cities or the cities and rural areas were not reported, but they apparently were not large, and for a number of items the adjacent rural areas had higher prices. This should not be surprising. As one moves to the center of urban areas commodity prices rise because of the higher site rents of retail locations, but there are often scale economies of distribution to market centers. This is not the case for services but pricing of services has been weak in all phases of the ICP including 2005. We first deal with the adjustments made to Chinese prices in the ICP, and then ask how different is the Chinese situation compared to other large countries.

Treatment of Chinese prices in the 2005 ICP

The expenditures refer to all of China so the Asian Development Bank (ADB) felt it necessary to move the urban prices to an all-China basis to replicate the inputs of fully participating countries. The relationship of urban to rural prices and of prices across regions is the critical step for such an extrapolation, especially since there have been no official studies of rural-urban or regional price differences. The ADB used cluster analysis on a number of characteristics of the 11 urban areas and other non-sampled areas of China. Most of the indicators entering the cluster analysis dealt with provincial incomes or correlates and no independent price information was considered even though some rough information was available. Without introducing some price information for rural China beyond that in the hinterlands of the 7 cities, it is not clear that any technique like cluster analysis can improve the situation.

Brandt and Holz (2006) have made the most comprehensive set of comparisons of rural-urban and regional price levels in China for 1990 updating the results to 2004. As more and more of the urban housing is market priced, the rural-urban differentials for rented and owner-occupied housing have increased in China. However, in their work Brandt and Holz only approximate rental differences by the cost of construction taking no account of the site rent of land. If there is a direction of error in their estimates, it is to

understate the difference between rural and urban prices in China. A common or joint basket of goods that holds quantities equal in 2004 showed more difference in prices across the provinces than between rural and urban areas within provinces. For example, the joint basket in rural Beijing is 84.7% higher than in rural Chongquing. The largest urban-rural difference within a province is 43.5% in Chongquing. The costs of a common basket in urban Beijing is 50.9% higher than in urban Chongquing. All of these differences in Brandt-Holz appear much larger than were used in the ICP and evidence support those who believe that the aggregate price levels in the 2005 ICP in China are too high.

The Brandt-Holz work is based on unit values and has other limitations that leads one to ask whether the differences they report by region are high relative to other countries. Aten (2006) reports that for the 38 urban centers used by the US for the CPI the differences between small southern urban areas and San Francisco are large, 80 versus 130% of the US average in 2003. From the million plus prices collected, Aten is able to obtain about 25,000 annual average price observations for 256 entry-level items collected by the BLS from which price level differences over all of consumption can be estimated. This is a rich data set that has now been updated to include 2004 and 2005 with similar findings, so that we can be fairly certain that the range across US urban areas is around 60%, suggesting that the Brandt-Holz estimates for China are not unreasonable. Aten also finds that the gradient of prices from low to high is not large for goods, but it is much steeper for services, a common finding of previous rounds of the ICP across countries. Unfortunately, it is service items like housing, medical, and personal services that have not been surveyed or measured very well in the ICP, or the expenditure surveys that underlie the Brandt-Holz study.

China and other large Developing Countries in the ICP

Is China different from other large countries in the ICP? First their sample of 11 large cities is a large number of people, but less, compared to the 38 centers for the US. But note that for the CPI in the US, there is no attempt to cover all states or urban areas for the reason that in the short-run prices move together over time. And in the US where catalogue and online shopping are common, and where rural inhabitants can buy big ticket or basic staple items at accessible malls, the CPI sample represents most purchases.

Again housing services are an important exception. In large developing countries like Brazil, China, India, Indonesia, or Nigeria there may be more frictions in the transmission of price shocks, but it is probably not necessary to sample beyond urban areas for purposes of capturing movements in time-to time indexes.

The problem for the ICP is the sampling frame for price collection in large and small countries. For Belize, Bermuda, Hong Kong, Luxembourg or Singapore, the frame of outlets for the CPI covers the country and thus represents a framework to build upon for the ICP. But for place-to-place comparisons, this is much less true for large countries. Not only are large parts of the countries not covered in the CPI, the existing outlet sample in less affluent countries is not well suited to the ICP list of items. As a consequence the degree to which large developing countries rely mainly on urban prices varies greatly across countries. India has a long tradition of collecting rural and small center prices, while Brazil, Indonesia, Pakistan and Thailand have mainly collected urban prices. However, even in India house rents are not estimated or imputed in rural areas in the price index for agricultural laborers. And the usual outlets for collection of rural prices, would not include outlets in smaller centers where some of the ICP items are more likely to be available.

Until there has been more analysis of the detailed ICP results not much more can be said about how we should interpret price levels between small and large developing countries as reported in the 2005 ICP.[4] One approach to this is to compare unit values from expenditure surveys in ICP countries, that do cover the whole of countries, with corresponding ICP prices. We do have estimates noted above that prices in rural China may be 35% below those in urban areas for a poverty bundle. For all of China, however, this affect for rural-urban price difference is likely to be less because for those not in

[4] Large high-income countries like the United States, France or Australia essentially provide urban prices for commodities and services. While this is similar to what is done in many of the lower income countries in the ICP, the consequences are not large because the higher income countries are more urban and make more purchases in urban areas or online. The EU asks countries to supply an adjustment factor to move urban prices to a national level, but the factor used for most items is 1.0 meaning no adjustment. There is a directive in the EU for Eurostat to estimate regional price levels within countries but this has not been funded or implemented. An important exception is housing, where EU countries collect rents on several sizes of apartments and houses, with different amenity combinations on a more national level. The US typically estimates a hedonic regression using the appropriate specifications to supply rent price levels.

poverty, a higher proportion of the purchases are for goods for which price differences are less. However, regional differences appear to be large in China, and it would appear that price collection took place in urban areas with higher than average prices ao there would be an added downward regional effect that also should be considered. It should be made clear that China has been very clear on where they would price for the 2005 comparisons, and the real problem is how to interpret the results. While these comments may sound critical of the 2005 benchmark, it is important to understand that we are only able to raise these questions because the 2005 comparison has been better documented by individual countries and by regions compared to the 1980 and subsequent benchmarks, when regional comparisons and fixity were introduced.

The Message for the 2011 ICP

A lesson learned is that alternatives to national average prices need to be explored in the next round of the ICP, which is preliminarily planned for 2011. A model for this was established Kolkoski, Moulton and Zieschang (1999) in their spatial analysis using the CPI checklist prices used for the CPI for 1987. An elaborated Country Product Dummy (CPD) equation was employed that not only used BLS regional center and product in the equation but other characteristics of the item, like weight and packaging, and of the outlet as price determining variables. Aten (2006) has resumed this analysis beginning with 2003, with prospective annual updates, using a stochastic method of aggregation, a weighted CPD or Rao approach. Future ICP rounds should seriously consider a hedonic type approach that could deal with both the representivity issue as well as national average prices.

Comparing Housing Services

The framework of earlier ICP comparisons was to use market rent comparisons for various size and amenity groups of housing and assume rental equivalence for owner occupied housing.[5] The EU and OECD countries used a similar approach until their expanding memberships included countries that were not suitable for surveys of market rents. A new member country might have a small expatriate community that paid market rents, and if other rentals existed, they were subsidized. The approach of the EU was to

[5] Some countries had large rural housing stocks that were not typically rented. Further these units are typically not comparable to urban rentals, so a more ad hoc user cost approach was used.

make direct comparisons of quality-adjusted volumes of housing in such countries and to find a link member country, initially Austria, or member countries that would both survey rents and provide quantity information on their housing stock.

For the 2005 comparison the plan was to use both a quantity and survey approach or some combination in other regions. In practice the quantity approach was used in South America, and a combination in Western Asia. In Asia, however, neither approach appeared feasible for all countries so the decision was to assume that the per capita volume of housing services was the same as the remainder of actual household consumption. The same approach was adopted in Africa. One consequence is that it is not meaningful to compare housing volumes in any country in Asia and Africa with a country in the other reasons. This is clearly an area that requires more work in the 2011 ICP not only to ensure that all regions follow a similar methodology but also to re-think the quantity approach. When quality adjusted quantities are compared across countries, it appears that they over-state the volume in lower income countries compared to higher income countries. The reason is that data from housing censuses have only a small number of amenities associated with any housing type so that many of the quality features associated with better housing do not enter into the estimate of the volume of quality adjusted dwellings.

The Equal Productivity Assumption

How does one compare the output of civil servants and health and education workers across countries? These comparison-resistant services also plague constant price estimates in the national accounts because the outputs are not typically priced. In national accounts deflation it is necessary to make some assumption about what is happening to the productivity of such workers over time, and the same is true across space for the ICP. In past ICP rounds volumes have been derived by dividing compensation by a PPP that was derived from a detailed comparison of salaries for specific occupations. It had been recognized that this procedure assumed equal productivity across countries in a given occupation, which was unlikely given very different amounts of accompanying capital per worker across countries. Further, there is much less inducement to organize the work environment to improve productivity of employees in administrative, health and education services in very low-wage economies.

In the 2005 benchmark, the range of countries was much greater than in previous rounds, and the consequences of the equal-productivity assumption loomed much larger. In Asia for example, salaries for the same occupation differ by a factor of 100 between Laos and Hong Kong. Similar differences exist between Yemen and Kuwait in the Western Asia comparison. Without some adjustment for productivity, the resulting per capita volumes of comparative-resistant services in Yemen or Viet Nam would greatly exceed those of its richer neighbors, an improbable outcome. Such adjustments have been considered earlier by the OECD and the ICP, but the 2005 Asian comparison is the first actual case where the equal productivity assumption has been significantly modified.[6]

Asia, West Asia and Africa have also carried out such adjustments based on estimates of capital per worker in the whole economy of each country. In Asia, for example, it means that the volume of GDP of China and India relative to Hong Kong or Singapore will be lower than in previous ICP rounds. This poses a problem of comparability across regions in 2005 because EU-OECD-CIS and South America did not make such adjustments. Further, because capital per worker data were not available for many countries, it was often necessary to apply the same adjustment factor to low-income countries that were at different stages of development. The actual procedure used is described in the final Reports of Asian Development Bank and World Bank.

Clearly this adjustment is in the right direction and earlier benchmarks did attribute too large a volume of such services to poorer countries, and consequently imparted an upward bias to their PPP converted GDPs. However, the particular procedure was based upon limited information applied uniformly over groups of countries within each region, so there is an unknown, but significant, error associated with the actual adjustments, even for countries within the same region. Further the

[6] The report on the 1975 ICP round (Kravis, Heston and Summers, 1982, p. 140) found that the price levels for non-priced services were lower than those of priced services, especially for low-income countries. However, it is difficult to substitute priced for un-priced services because most countries do not collect an adequate number of prices for purchased services, but it would be an improvement on the equal productivity assumption. In addition an attempt was made to obtain capital stocks used in education facilities to adjust for productivity, but country response was spotty so the experiment was limited.

adjustments in Africa and West Asia were each calibrated differently than for Asian countries. What does this mean for comparing the 2005 results to previous benchmarks? In previous benchmarks, the volume of administrative, health and education services for very low wage countries in Africa, Asia, and W. Asia would have been substantially lowered if the 2005 procedure had been adopted in those years. Everything else the same the methods adopted for these sectors has the effect of producing a smaller spread in real GDP per capita between rich and poor in 2005 than in previous benchmarks. Because of the increased spread in the 2005 ICP, some observers have concluded that globalization has increased world inequality. Because of the change in method, that inference is not justified.

What is the consequence for the 2005 comparison of the mixed application of an adjustment for productivity in some regions and not in others? Certainly Asian GDP was reduced compared to the OECD countries as a consequence of the productivity adjustment.[7] This means that comparisons of Asian countries with peer countries like Brazil, Mexico and many eastern European countries where the productivity adjustment was not carried out, would also be affected. This is not an argument against a productivity adjustment, though the actual implementation was of a 'one size fits all' countries nature; rather it helps us better understand where the new view of the position of China and India in the Global economy of 2005 is coming from. And it certainly points to the need to gain agreement on a standard method of treating un-priced services among all the regions.

Linking of the Regions

In previous global comparisons linking of regions has often been through only one or two countries, in which case the results can be quite sensitive to the particular link countries.[8] In 2005 a method that was less sensitive to the choice of countries was

[7] Asia does not include Korea and Japan in this case as they are treated in the OECD in the Global Regions.

[8] Also the linking can be done at a detailed level or an aggregate level. When it is done at an aggregate level as in 1985, it is particularly sensitive to the link countries, e.g., Japan was used to link Asia to the OECD.

adopted for linking regions at the basic heading in the ICP.[9] These basic heading parities in each region were used to convert the national currency expenditures in each country to a volume in the currency of the numeraire country of a region, like Oman in Western Asia. The next step is to aggregate these expenditures and parities for each region to a total, like consumption or GDP. In the 2005 this was done in a novel way. First, the basic heading volumes in the numeraire country of each region were added up over countries for each basic heading to give regional totals. Regional price levels can then be derived for each basic heading and aggregations were carried out by region.[10] The advantage of this approach is that it preserves the relationships of countries within a region for each basic heading thereby maintaining fixity. But this is also the disadvantage of the approach as applied in 2005.

The method of aggregation used was the EKS method, which in effect gives equal weight to Africa, Asia, OECD, South America and Western Asia.[11] The building blocks of EKS are all possible binary Fisher indexes. Table 3 presents in the upper diagonal the Paasche-Laspeyere (P-L) spreads for all possible binary comparisons for six regions. In fact, in the calculations, CIS was included with the OECD, so that only 5 regions were used in the EKS. This is important to note because the direct comparisons between any

[9] In the parlance of the ICP there were 18 Ring countries from the 5 regions that undertook special pricing. Based on prices from these countries parities at the basic heading level were estimated for the Ring countries that could be linked to each of the 5 regions. The method is described in Diewert (2006). The method converts the prices of each Ring country in a region to a price expressed relative to the regional average, so essentially the price reflects information about the Region.

[10] The regional basic heading price level is derived by dividing the regional total for a basic heading at exchange rates by the total at PPPs all in the numeraire currency of the region, say the HK dollar for Asia. Note that CIS is presented as a region; it was linked to the OECD through Russia.

[11] The method of Elteto, Koves and Szulc (EKS) as well as the method used in PWT, namely the Geary-Khamis (GK) average price approach are described in Kravis, Heston and Summers (1982, pp.88-9) and in the ICP handbook on the World Bank ICP site. EKS, whose origins trace to Gini (1931, p.12), works with binary comparisons between each pair of countries that are not transitive; that is the binaries (A/B) / (A/C) will not in general equal the direct binary C/B. The EKS algorithm smooths the direct and indirect comparisons of each country so that final relationships will be transitive. GK is also a transitive method. It achieves this result by deriving a common set of international prices with which to value quantities in each country. GK is not a superlative index in that country quantities are not allowed to adjust to the international prices. Both methods are base-country invariant. GK is additive, though not EKS. The African region wanted additive results, and used the Ikle method as made operational by Dikhanov (2005).

two regions only receive 40% of the weight but each indirect comparison receives 20% when there are 5 regions.[12]

Table 3: Paasche-Laspeyere Ratios in Upper and Price Similarity Indexes in Lower Diagonal

Region	Asia	LAC	Africa	OECD	CIS	W.Asia
Asia	1.00	1.28	2.06	1.60	1.74	1.34
LAC	.935	1.00	1.42	1.62	1.71	1.42
Africa	.922	.938	1.00	3.00	3.17	2.13
OECD	.824	.867	.790	1.00	1.07	1.76
CIS	.794	.847	.764	.979	1.00	1.67
W.Asia	.898	.929	.891	.857	.857	1.00

Note: This illustration is based on an earlier version of the final data set. The price similarity measure is the correlation between the relative price vector of each region weighted by the average expenditure share of each pair of regions.

Except for the CIS region, the P-L spread is large between most regions and the OECD, especially Africa. This means that the possible error in the EKS estimate is also likely to be significant. The lower diagonal of Table 3 provides a price similarity index

[12] That is, the Asia to LAC relationship would be counted twice and the three indirect relationships through Africa, Asia and OECD each once in the EKS calculation.

between the regions, which is constrained to be between 0 and 1.[13] Usually low price similarity and high P-L spreads go together. The first point to be made here is that by treating the region as one unit in the world aggregation, it appears very little is gained in terms of reducing differences in price structure over the aggregation. The second point is likely to be more important. Maintaining the price structure within each region means that the resulting global comparisons reduce comparability between any pair of countries in two different regions. For aggregations like food, transport or investment as well as aggregate GDP, comparisons of say, Brazil and India, or China and Russia are not free. The protocol to allow researchers to have access to the basic data is not yet in place, so it is not yet possible to illustrate the quantitative importance of our conjecture. These linking issues arise because of the fixity requirement. However, for global comparisons the importance of comparisons within regions has certainly diminished relative to the importance to comparisons of economies across regions. Smaller countries in Africa, Asia or South America are will be mainly interested in their relation with peers in their region, but this is much less the case with the larger countries whose peers are often in other regions.

One final observation on aggregation procedures used in the ICP. In PWT it has been the practice to use the Geary-Khamis (G-K) method, which was also used in the global comparisons in 1970 and 1975. The OECD also publishes a version of G-K one year after they release their benchmark EKS results. If aggregation of the 5 regions were carried out using the G-K method it would, in fact. be little different from the EKS results. The reason for this is that the average prices used in G-K are now nearer the average for the world than in earlier benchmarks. This is also true if G-K were run on individual countries, though in the latter case individual countries would appear quite different than the 5 region aggregation,

B. Extensions of the Benchmark Estimates Over Time and Space

Since the ICP was begun in 1968, there has been an interest in covering non-benchmark countries, and PWT was launched with this in mind. In addition the

[13] The price similarity measure used is discussed in Heston, Aten and Summers (2) In the 1975 ICP involving 34 countries, most values of the 561 similarity measures were between .85 and .90, with a few as low as .67 and many, like France and Belgium at .99.

international agencies also have an interest in covering all their member countries. A brief discussion of non-benchmark estimates is given in the first part of this section. The second section takes up one issue of extending benchmark estimates backward and forward over time, namely foreign trade. The third section focuses on the extensions over time in PWT, and the important lessons that can be taken from recent research undertaken at IMF.

Non-benchmark Estimates

When non-benchmark estimates were launched in the mid 1970s the number of benchmark countries was only 16. In 2005 the task is much less daunting because there are 146 benchmark countries, so the number of non-bench mark countries is under 40, many in the Caribbean because they were the only significant country grouping that did not participate. The Bank did ask consultants to suggest how do estimate non-benchmark countries for 2005 but were not able to use any of their suggestions. As a consequence they used the estimating equation that had been developed for WDI estimates for a decade or more:

(1) $PPP_{GNI} = \beta_1 \text{Atlas} + \beta_2 \text{SecEd}$,

where the PPP for Gross National Income (GDP + net income from abroad) is regressed against GNI per capita based on the Bank Atlas averaging of exchange rate converted GNI and Gross High School Enrollment Rates.

The two main differences between the Bank approach and PWT is that the latter works with Domestic Absorption (DA), using the log of PL_{DA} as the dependent variable and takes advantage of some price information, however imperfect, on the right hand side. The price information is in the form of post adjustment indexes of the International Civil Service Commission (ICSC), and the US and Canadian foreign offices. In addition we use openness to trade in the form suggested by Kravis and Lipsey (199), namely that the price level rises with openness but with a dampening effect as GDP per capita rises. And we introduce a new variable that classifies countries into high, medium and low involvement in international financial flows as measured by the Bank for International Settlements.[14] This variable labeled FF, is an attempt to capture the unusual situation of

[14] The series used was '17_Liabilities to BIS banks (cons,) short term'. Countries were divided into 5 groups by the ratio of these liabilities to GDP, and later reduced to 3 groups.

some of the off-shore financial centers. Finally we introduced regional dummies for Africa and the OECD, the other regions not being different from each other. The result with standard errors in ()s is:

(2) $LnPL_{DA}$ = -1.299 + .665 CANPA - .025 USPA - .218 ICSCPA +.457 ln OPEN
(.169) (.166) (.245) (.260) (.260)

-.620 ln(OPEN*y) +.351 FFHIGH + .173 FFMED +.316 Africa + .139 OECD
(.169) (.076) (.067) (.066) (.063)

Only 74 of the 146 countries in 2005 had all 3 post adjustment indexes, which are naturally highly correlated with each other. As a consequence only one, the Canadian index turns out to be significant. However, we want to be able to estimate non-benchmark countries that may have one or more of the 3 indexes. In terms of previous equations in earlier versions of PWT, (2) performs well, with an adjusted R^2 of .904 and more importantly a RMSE of .160, whereas estimating equations using earlier benchmarks typically had RMSEs of over .220. When the secondary education measure of the Bank is introduced into (2) it is not significant. In fact work within the Bank subsequent to the Final Report suggests that there are other variables that might be added to improve on (1) above, that the equation is improved by using a semi-log form, and that education does not then enter the equation.[15]

The reason that PWT uses domestic absorption is that handling of trade is weak in both the benchmark comparisons as well as in extrapolations over time. Turning to the latter point first, the WDI for example relies on extrapolations for GDP, and certainly this is one reason that their estimates for 2005, which relied on extrapolation of 1993

[15]These equations regress log of per capita GDP from the 2005 benchmark expressed relative to the US on exchange rate converted GDP and a variety of other variables. In this form R^2 values look very high (i.e., .98-.99) because of the wide country variations in per capita GDP, so a better measure of goodness of fit is the mean square error. In the short-cut estimates used in PWT the log of the price level of Domestic Absorption ($100*PPP_{DA}/ER$) is used as the dependent variable, which has much less variation across countries and less inflated R^2 values. For either form, the mean square error is the better goodness of fit measure. When additional variables are added, the secondary education variable is not significant, which is no loss since it was difficult to interpret the sign or rationale for the variable. In fact, their preferred equation appears to contain dummies for each region and for island economies. In terms of comparison with what is done in PWT, the RMSE in this equation is roughly .185, which may be compared favorably with a RMSE in the actual equation used by the Bank of .225.

benchmarks, were often quite different than the new benchmark estimates. This means that if export volumes are constant but their prices fall, as in the case of micro-chips for Singapore, GDP growth will overstate the ability of Singapore to convert current production into current domestic expenditures in 2005. Other changes in the terms of trade will similarly drive a positive or negative wedge between extrapolations and current price PPP conversions.

As argued in Feenstra, et.al. (2004) PWT and the ICP really provide an income as opposed to output estimate in any year. The net foreign balance in PWT is converted at the PPP for domestic absorption, which has the virtue that it is base country invariant. When the net foreign balance is converted at exchange rates as in the 2005 ICP it treats countries differently depending on whether the balance is positive or negative and their price level is greater or less than 1. For example, the PPP for both China and India is less than their exchange rate. China's surplus is therefore converted at a value less than its command over goods in the Chinese market, while India's deficit is converted to be a smaller reduction than it would be at domestic currency. However, the way to obtain real output is to convert exports and imports at their PPPs, which was the exercise that was implemented using unit values for 1996 in Feenstra, et.al. (2004). When this was done, significant differences were found between output and income suggesting this is a fruitful line of research to pursue, notwithstanding the difficulties of disentangling quality from unit value differences.

Extending PPP Estimates Backward and Forward in Time

The previous section discussed the PWT approach to disaggregating the extension of benchmark PPP estimates over time and space. Distinguishing between the foreign sector and domestic absorption is important, but it is not clear that the way we have actually computed real GDP over time in PWT was warranted. Let us start with the current price estimates in PWT. The PPPs for C, I and G for a base year, say 2000, are moved backwards and forwards in time by the deflators of each aggregate relative to the US. Then a new aggregation is carried out for say, 1995, that provides a current price estimate of the 1995 PPP for DA. The net foreign balance in current prices was converted at the PPP for DA to obtain GDP. This provides a time series in current prices.

The questions arise with GDP in say 2000 prices, especially because this is the most commonly employed series by users in their models. PWT has offered two principal constant price measures, a fixed weight and chain weight index. The fixed weight index in PWT 6.2 uses the share of C, I, and G in 2000 as the weights applied to national growth rates for each of these expenditure aggregates. The chain weight index applies the current price weights of the year *t* to the growth of C, I and G between *t and* $t\text{-}1$. This provides an estimate of the growth rate of DA between *t and* $t\text{-}1$ to apply to the DA in year t in 2000 prices. Many inquiries have been fielded about PWT for the past 25 years but none have questioned the fixed and chain indexes, which should best be interpreted as benign neglect rather than critical acceptance. However, for a chain index, given the limited national accounts detail available, this is the only alternative. For a fixed weight index it would be possible to simply use the national growth rate of Domestic Absorption, and then add the net foreign balance, a series that will be provided in the future.

How should the NFB be converted in constant prices? In previous PWTs the growth rates of exports and imports was applied to the 2000 values of exports and imports for each country converted at the PPP for DA. An alternative treatment was also offered that attempted to take account of the terms of trade. Neither method is satisfactory. National growth rates of exports and imports usually reflect changes in production with fixed weights. Trade is an important area for improvement in both the ICP and PWT.

Is PWT Consistent across Versions?

As PPP estimates have evolved over the years the underlying data-base of PWT has also been revised. New benchmark estimates bring in additional countries and revised estimates for multiple benchmark countries. National accounts are subject to revisions, and changing base years of PWT also introduces elements of non-comparability between different versions of PWT. Users have been advised of these changes with each updating of PWT, but has the advice been heeded? Happily ongoing work at IMF undertaken by Simon Johnson, William Larson, Chris Papageorgiou and Arvind Subramanian (JLPS) sheds some light on this question. At this stage of their

work the authors would prefer that the specifics of their results not be quoted, so the following is a general summary of their findings thus far.

Differences in Growth Rates between PWT versions

Annual growth rates of GDP based on the chain series were compared from PWT 6.1 and 6.2 for 40, 10 and annual intervals. Why would they differ? The discussion above has suggested the main reasons: weights for growth of C, I and G change for both the fixed and chain indexes, national accounts are revised and rebased, and new benchmark information becomes available. For all countries the annual cross section is fairly similar, the ten-year growth rates generally differ by under10%, but the 40 year rates often differ by 20% with some embarrassing outliers. Differences are lower for high-income countries, and larger for both low and middle-income countries. Of more relevance is the standard deviation of the growth rates. For the 1 year panel of 5353 observations the sd of the differences is .0625, for the 10 year panel of 508 observations the sd is .0163 and for the cross section of 151 countries it is .0127. Given that the average growth rate in PWT 6.2 is 0.019, the sds are large even for the cross section.

In PWT quality grades have been assigned to countries on an A to D standard based upon number of benchmark comparisons in which a country has participated and some internal measures of data stability over time. With the exception of Singapore, there is for practical purposes no difference in growth rates for the remaining 30 A and B grade countries.[16] If users have employed PWT grades as variables or to group countries in their analysis, it has not come to our attention. In work with earlier benchmarks it was found that differences in growth rates in PWT and in national growth rates were largest for countries where their own national growth rates were affected by using a new base year.

[16] As suggested earlier, the case of Singapore is instructive. PWT 6.2 uses a 2000 reference year at which time the unit price of Singapore's electronic exports had substantially declined compared to 1996, the reference year for PWT 6.1. This substantially lowered the weight applied to the growth of volume of electronic exports, and to a lesser extent the GDP growth of Singapore between the two versions of PWT.

Replication Studies and Associated Warning Labels

Do these growth differences between versions of PWT make a difference? SLPS undertook an extensive literature search and performed a number of replications involving PWT 5.6, PWT 6.1 and PWT 6.2. The preliminary guidelines thus far are:

a. SAFE: Studies that mainly use long-term growth, 40 year intervals, are fairly robust with respect to use of any of the above versions of PWT. This may also be true for 10 year intervals, or at least the conclusions of such studies do not appear dependent on which version of PWT is used.
b. SAFE: Use of annual growth rates for the A and B grade countries is safe.
c. It is not safe to use annual growth rates from different versions of PWT for non-OECD countries.
d. NOT SAFE: Different versions of PWT are not robust for analysis of turning points, or year-to-year movements.

Conclusions

The IMF group plans to fine-tune their analysis and to see if PWT 7.0, which will incorporate the 2005 ICP, leads them to change their story. This brings us full circle to the problem of integrating the new view of the world economy in the 2005 ICP with the older view in the WDI, WEO and PWT. As in the past, the plan is to present what is a reasonable view of the world economy in 2005 and to move that backward and forward in time. The work of SLPS suggests that PWT should follow national statistical practice and provide the old world in on a 2005 base, PWT 6.3; and PWT 7.0, which will incorporate the 2005 ICP. The SLPS research also suggests that more alternatives need to be considered for moving the PWT numbers over time. As presently constructed and updated PWT provides a panel of annual data, but the annual data need to be used with caution.

Would the results of ICP 2005 be woven into PWT 7.0 without adjustment? The answer is not before investigating several possible adjustments including how different would be the results without the fixity constraint. Other possible adjustments have also been discussed above including modifications for the special character of Chinese prices, the lack of comparability of non-priced goods and services across the regions, and the very special way in which relationships within regions were fixed in the 2005 round.

Until access to the underlying parities and expenditures is resolved, it is not possible to judge the impact of these adjustments on the 2005 benchmark. What can be said is that because of ICP 2005 there is a much richer data set available for those researchers interested in differences of economic structure and income across countries than has been available until now.

References

Aten, Bettina 2006. 'Interarea price levels: an experimental methodology', Monthly Labor Review, September

Brandt, Loren, and Carsten Holz (2004) "Spatial Price Differences in China: Estimates and Implications," Economic Development and Cultural Change

Diewert, Erwin (2006) "On the Stochastic Approach to Linking the Regions in the ICP," http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/ICPEXT/0

Dikhanov, Yuri (1997), "Sensitivity of PPP based Income Estimates to Choice of Aggregation Procedure," http://siteresources.worldbank.org/ICPINT/Resources/icppapertotal.pdf

Feenstra, Robert, Alan Heston, Marcel P. Timmer and Haiyan Deng "Estimating Real Production *and* Expenditures Across Nations: A Proposal for Improving the Penn World Tables", with. NBER Discussion Paper # 10866, Oct 2004. Revision accepted in Review of Economics and Statistics. In Press.

C. Gini (1931), "The Circular Test of Index Numbers," Metron: International Review of Statistics, Vol IX-N. 2.

Johnson, Simon, William Larson, Chris Papageorgiou and Arvind Subramanian (2008) , "Some Health Warnings About Using PWT Growth Rates," Presentation. PWT Workshop, May 11th.

Kolkoski, Mary, Brent Moulton, and Kim Zieschang (1999) "Interarea Price Comparisons for Heterogeneous Goods and Several Levels of Commodity Aggregation," in International and Interarea Comparisons of Income, Output and Prices, ed., by AlanHeston and Robert Lipsey, 123-66. University of Chicago Press.

Kravis, I.B., A. Heston, and R. Summers (1982), World Product and Income, Baltimore: Johns Hopkins University Press.

Kravis, I.B., and Robert Lipsey (

Maddison, Angus and Harry Wu (2007), 'China's Economic Performance: How Fast has GDP Grown? How Big is it Compared with the USA?, Paper Presented at the

International Conference on Experiences and Challenges in Measuring National Income and Wealth in Transition, September 18-21, 2007, Beijing, China, September

Maddison, Angus (2007) Chinese Economic Performance in the Long Run, Second Edition, Revised and Updated, OECD, Paris.

United Nations Economic and Social Commission for Asia and the Pacific (1999) ESCAP Comparisons of Real Gross Domestic Product and Purchasing Power Parities, 1993, (ST/ESCAP/1918), Bangkok