

Understanding PPPs and PPP-based national accounts

Angus Deaton, Research Program in Development Studies, Princeton University
Alan Heston, University of Pennsylvania

November 2009

We are grateful to Shaida Baidee, Misha Belkindas, Erwin Diewert, Yuri Dikhanov, Nada Hamadeh, Mick Silver, Eric Swanson, Fred Vogel, and two referees for helpful comments on an earlier version. Both authors were members of the Technical Advisory Group for the 2005 International Comparison Program, which was chaired by Heston. This paper represents their personal views, and should not be taken to represent the views of the ICP, the World Bank, or any other international institution, none of which supported this work financially. Support from NSF grant SES-0648769-002 is gratefully acknowledged by Heston. Parts of this paper were presented at CRIW-NBER Summer Workshop, July 14-15, 2008.

Understanding PPPs and PPP-based national accounts

Angus Deaton, Research Program in Development Studies, Princeton University
Alan Heston, University of Pennsylvania

November, 2009

ABSTRACT

PPP-based national accounts have become an important part of the database for macroeconomists, development economists, and economic historians. Frequently used global data come from the Penn World Table (PWT) and the World Bank's World Development Indicators; a substantial fraction of the world is also covered in the PPP accounts produced by the OECD and the European Union. This paper provides an overview of how these data are constructed, and discusses both the theory and the practical problems of implementing it. All of these data are underpinned by the International Comparison Program (ICP), which collects data on prices worldwide. The most recent round of the ICP was for 2005 with final results published in early 2008; version 7.0 of the Penn World Table will soon incorporate these results. The 2005 ICP, like earlier rounds, involved substantial revisions to previous data, most notably revising downwards the size of the Chinese (40 percent smaller) and Indian (36 percent) economies. We discuss the reasons for the revisions, and assess their plausibility. We focus on four important areas: how to handle international differences in quality, the treatment of urban and rural areas of large countries such as China, India, and Brazil, how to estimate prices for government services, health, and education, and the effects of the regional structure of the ICP. All of these affect the interpretation of previous data, as well as the current revisions. We discuss previous revisions of the PWT, and their effects on various kinds of econometric analysis. The paper concludes with health warnings that should be kept in mind when using these data, which are not always suitable for the purposes to which they are put. Some international comparisons are close to impossible, even in theory, and in others, the practical difficulties make comparison exceedingly hazardous.

1. Introduction

Over the last thirty years, more and better data have fundamentally changed the practice of both microeconomics and macroeconomics. No new data have been more important and more influential than those from the International Comparison Program (ICP). The ICP collects prices in countries around the world, and uses them to calculate price index numbers or purchasing power parities (PPP), whose aim to measure how much local currency is needed to buy as much as does the currency in the numeraire country, usually the US dollar. The “as much” can refer to GDP, or to one of its components, such as investment or consumption. As with price indexes within a country, PPPs can be thought of as statistical averages of prices, or given a cost-of-living interpretation. They are also used to deflate nominal local currency measures to yield “volume” measures expressed in a common currency unit, such as current US dollars for the year of the comparison. Adjusted for inflation in the numeraire country, the ICP yields real GDP accounts in constant internationally comparable dollars.

By the late 1960s, the theory of economic growth that had begun with Robert Solow’s great paper had become a largely theoretical enterprise. But by the late 1980s, the Penn World Table had evolved from a small set of illustrative calculations into a multi-country panel big enough for econometric analysis, particularly Mark 5 which contained up to 39 years of data on 138 countries, Robert Summers and Alan Heston (1991). These data helped bring about a new growth economics, with theoretical developments consistently related to evidence. There has been a huge explosion of work since then, trying to understand the mechanics of growth, linking growth and politics, and forging an integration of macroeconomics, economic development, and economic history, the last

supported by the companion creation of long-run historical data by Angus Maddison (2003). Purchasing power parity data provide a common measuring rod that allows comparison, not only of India and America now, but of India now with Britain before the industrial revolution. It allows the World Bank to estimate the number of poor in the world, and permits plausible conjectures about which places and which times have seen the greatest riches and the greatest poverty in human history, Lant Pritchett (1997).

Although the international data are widely used, the way that they are constructed is not always as widely understood. Nor is it easy to find out why different, commonly used sources give different estimates. Perhaps the most familiar of these sources are the PPP data given in the PWT, on the one hand, and in the World Bank's World Development Indicators (WDI) on the other. Eurostat and the OECD currently cover 55 countries in the OECD, Europe, and the Confederation of Independent States (CIS), and publish time-series data back to 1980 for a subset of those countries. This paper aims to provide an overview of the most recent round of data collection, together with the underlying economic theory, an explanation of why different sources give different numbers, and some health warnings for their use. We pay particular attention to the PWT because it is the only source that gives long time series for a disaggregated set of national accounts. There is an enormous amount of detail that goes into the collection of the data and the construction of the accounts, but our aim here is to focus on a few key issues that are likely to be important to practicing macroeconomists, and where we think more knowledge is likely to be useful in practice.

An overview of the most recent round of international price comparisons is provided in the final report of the ICP, World Bank (2008a), and the details can be found online in

the ICP Handbook, World Bank (2008b). Our account here starts with the price-index theory that underlies the international accounts, and emphasizes the differences between multilateral and standard price comparisons. This also allows us to explain why the same set of underlying prices can give rise to several sets of apparently inconsistent national accounts. We then discuss some of the practical difficulties. Among the most important of these are how to handle international differences in the quality of goods and services across countries, how to price “comparison resistant” items such as government services or the imputed rents of owner occupiers, and how to calculate accounts for countries and periods that are not covered by the underlying data collection. There are also specific “health warnings” with any specific set of accounts; in the latest round the most important concern the regional structure of the data collection, the balance between urban and rural data collection, and the role of India, China, and the countries of the former Soviet Union. Even so, we emphasize the great progress in the ICP over time; there are many fewer health warnings for the 2005 round than for earlier rounds, and many of the most marked differences reflect corrections of problems in earlier rounds. There are also important warnings that apply to all rounds, particularly the unsuitability of these data for analysis at annual frequencies.

Most economists who use PPP data would currently work with the latest versions of the Penn World Table, Versions 6.2 or 6.3, which do not include the data from the latest ICP round for 2005. Version 7.0 of the PWT, which is currently in preparation, will incorporate these data, and will adjust the past data to provide a coherent set of numbers as far back as 1950. So we emphasize some of the areas where the 2005 round is different from earlier rounds, in part because these will cause substantial revisions compared with

Version 6, but also to illustrate the changes that have taken place in the past with every new round of data collection. These revisions are often substantial. For example, the 2007 version of the World Development Indicators (WDI), World Bank (2007), lists 2005 per capita GDP for China as \$6,757 and for India as \$3,452, both in current international dollars. The 2008 version, World Bank (2008c), which includes the new ICP data, gives, for the same year, and the same concept \$4,088 for China and \$2,222 for India. For comparison, GDP per capita at market exchange rates is \$1,721 for China and \$797 for India. We shall have something to say about what drives these revisions but it is hard not to speculate about which previously established econometric results survive the incorporation of these revisions into the PWT.

Economists are most familiar with PPP accounts through the PWT and, secondarily, through Maddison's data. The underlying data all come from one or more rounds of the ICP, which started as a joint project between the United Nations Statistical Office (UNSO) and the University of Pennsylvania, then a UNSO project, and most recently—from the 1993 round on—a World Bank managed project endorsed by the UN Statistical Commission. The main business of the ICP is to collect data on the prices of thousands of comparable goods and services in many countries, 146 in the 2005 round. These prices, together with the national accounts for each country, are used to construct a set of price indexes that compare, for example, the price of consumption or investment in India relative to the price of consumption or investment in the US, expressed in rupees per dollar. “Volume” estimates—in temporally current prices—come from dividing expenditures by the price indexes, in the example, giving estimates of both Indian and US

consumption or investment in current international dollars. Several preliminary points follow from this structure.

First, there are many different reasonable formulas for price indexes, and these will give rise to different indexes and different sets of accounts, which is (one of the reasons) why the GDP estimates published by the World Bank in the WDI are different from those that appear in the PWT. Which of these is most appropriate depends on the purposes to which the data are to be put; as is the case with most index number questions, there is typically no unique right answer. These issues are familiar from standard within-country accounting where, for example, Paasche and Laspeyres indexes are not identical. But the differences tend to be more important in international comparisons over space than in national comparisons over time because the patterns of relative prices and of expenditures are much more different between India and the US, for example, than within the US or within India a few years apart.

Second, the ICP must typically rely on local country estimates of GDP and its components based on the production, expenditure or (in some cases) the income side of the national income accounts. Many non-OECD countries do not routinely disaggregate their GDP into the categories that are required for the ICP. An important goal of recent rounds of the ICP has been to improve statistical capacity in participating countries, which in 2005 included major efforts in developing more detailed and accurate expenditure estimates. The national income accounts of many low-income countries remain very weak, with procedures that have sometimes not been updated for decades, see in particular Dudley Seers (1983, 18–27) who describes many of the problem areas,

albeit as they were 25 years ago. In many cases, the prices collected under the auspices of the ICP may be more accurate than the GDP numbers with which they are combined.

Third, it is important to understand that aspects of the exercise are close to being impossible in theory, and are therefore not amenable to data improvement. Making price comparisons between Canada and the US, or between the countries of the OECD, is relatively straightforward because the same goods and services are widely available in all. But when it comes to comparing a rural Thai agricultural laborer, who lives on rice, with his Ethiopian counterpart, who lives on *teff*, we have no basis for comparison. Rice is hard to find in Ethiopia and *teff* is impossible to find in Thailand, so price comparisons are not possible. This is an extreme case, but many goods and services that are widely consumed in rich countries are not available at all in poor countries, or are only available at high-priced specialty stores in a few large cities. One general rule is that the comparisons become less reliable the further apart are the structures of GDP (or its components) of the countries being compared. This is essentially the same phenomenon as the increasing unreliability of long-run historical comparisons the further back we go.

With these caveats in mind, we plunge into the formulas, which assume that prices and expenditures are available for all countries on a common set of goods and services.

We shall return to the reality in Section 3.

2. Multilateral price indexes in theory

In a world in which the law of one price were true, market exchange rates would be all that we would need for converting accounts in one currency into another. The price of any item in one country would be the price in any other converted at the exchange rate,

and the same would be true for a price index for consumption, investment, GDP or whatever. For various reasons, see Kenneth Rogoff (1996) for a review, relative prices are different in different countries, so that it is useful to compare prices directly, and to calculate price indexes for GDP and its components. Given a set of prices of all the goods and services in consumption (say, and we use consumption as an example), it is straightforward to use standard formulas (Paasche, Laspeyres, or Fisher, for example) to compute consumer price indexes for any one country in terms of any other. But these “binary” indexes do not give us what we want. If there are M countries, they yield an M by M matrix of price indexes, not a vector of M price indexes, one for each country which, like exchange rates, would allow us a unique way of converting the price level of one country into another. More formally, we require that the matrix of the price indexes have two properties, first that the price index of country c in terms of country d should be the reciprocal of the price index of d in terms of c , and second, that the indexes be transitive, so that the price of Thailand with Botswana as base is the same whether computed in one step, or computed in two or many steps, from Thailand to Peru and then from Peru to Botswana, or through other intermediate countries. It is straightforward to show that these requirements are satisfied if, and only if, there exists a vector of M price indexes such that element c, d of the matrix is given by the ratio of the price index for d to that of c .

Perhaps the obvious approach to economists—though not to national income statisticians—is to work with cost-of-living indexes which, in theory, provide the price indexes that we need. If prices in country c are written as an n -vector p^c , and if tastes are homothetic and identical in all countries, the cost or expenditure function can be written

as $u^c \alpha(p^c)$ for utility level u^c and some linearly homogeneous function $\alpha(\cdot)$ which is not indexed on c . The $\alpha(\cdot)$ functions immediately give us the price indexes that we need, so that if we (arbitrarily) take country 1 as numeraire (the US in all of these calculations), the consumption PPP for c in international dollars is simply

$$P^c = \alpha(p^c) / \alpha(p^1). \quad (1)$$

These PPPs make no assumption about relative prices being the same in all countries; essentially the assumption of identical homothetic tastes replaces identical relative prices in allowing us to construct indexes. The indexes in (1) can be estimated by specifying a demand system and fitting it to the world data, or through a finite, nonparametric revealed preference approach pioneered by Sydney Afriat (1967), more recently developed by Steve Dowrick and John Quiggin (1994).

The assumption that all countries have identical homothetic tastes is contradicted by much empirical evidence. For example, it implies that the income elasticities of all goods are unity so that, at fixed relative prices, all expansion paths are rays from the origin, a proposition that has been falsified by more than 150 years of empirical demand analysis. That tastes are identical across countries implies that *all* differences in the patterns of demand must be attributable to international differences in the structure of relative prices. While relative prices are certainly important, so are habits and tastes generated by past consumption, David Atkin (2009), and there are many places in the world, such as North and South India, where there are large differences in consumption patterns of food in spite of only modest differences in relative prices, possibly because relative prices were much more different in the past. For a broad classification, such as food, water, and shelter, it might be argued that everyone's needs are much the same, but the ICP works

with 120 basic heads of consumption—basic heads are the level below which there are no expenditure data—and collects prices on more than a thousand individual consumption items. To assume that tastes are identical and homothetic when they visibly are not would compromise both the accuracy and the relevance of the calculations. For example, if we think about comparing India and China using assumed homothetic tastes that might be dominated by American consumption patterns (although inconsistent with them), it is easy to understand Afriat's (1972) claim that to make such an assumption leaves "the significance of such calculations quite obscure, even as to the locus of injustice."

Although identical *non*-homothetic tastes are usually assumed in trade theory, we would argue this is no more plausible as a description of the world. If tastes are identical but non-homothetic, then the COLI-based system of PPPs depends on a reference utility level—effectively real income—and this modification is not insubstantial. For example, it is not clear that it makes sense to compare Mexican and Peruvian prices as if both were as rich as Japan, or even at some mean level of world income. Peter Neary's (2004) GAIA system of PPPs, which is the leading example of the approach, is constructed on this basis, with explicit assumptions about tastes, and a common system of demand functions estimated worldwide. Identical tastes can be further modified to include "taste-shifters," such as temperature or rainfall, but this just extends the number of reference characteristics that need to be fixed and further stretches the credibility of the numbers. It seems odd, to say the least, to compare the relative costs of living in, say, Congo and Ghana under the supposition that both have the Russian climate. Indeed, many of us would argue that price indexes are not always the same thing as the cost-of-living indexes. If all prices were identical in Moscow and in Ouagadougou, it seems meaningful

to say that the *price level* is the same in both, even if the *cost-of-living* is higher in the colder, northern city. In the rest of this paper, we shall follow the national income accountants' practice of thinking in terms of price indexes, not cost-of-living indexes. Critics of this approach argue that it leaves the welfare basis of the calculations unclear; we agree, but do not know how to do better, and therefore endorse a statistical over a welfare interpretation of the price indexes from the ICP.

One element of cost-of-living theory remains useful in the international context; this is W. Erwin Diewert's (1976) concept of a superlative price index. Superlative indexes, such as the Fisher ideal index, or the Törnqvist index, are defined as indexes that are consistent with preferences that are flexible enough to provide a local second-order approximation to arbitrary preferences. Such indexes, unlike the Paasche and Laspeyres, construct indexes using weights that depend on both the reference and comparison situations, and automatically satisfy the reversal property, that the price level in d based on c is the reciprocal of the price level of c based on d . Diewert (2001) also shows that superlative indexes can be regarded as "symmetric means" of the two different indexes that we would otherwise have, just as the Fisher index is a symmetric average (here geometric mean) of the Laspeyres for d based on c , and the reciprocal of the Laspeyres for c based on d , which is identical to the Paasche for d based on c . Since we shall use the Fisher as one of our running examples, these relations are worth recording, and also allow us to establish some notation. With M countries, labeled c, d , etc, and N goods, labeled i, j, k , etc, the Laspeyres, Paasche and Fisher for d with c as base are

$$P_L^{cd} = \sum_{i=1}^N s_i^c \frac{p_i^d}{p_i^c}; \quad P_P^{cd} = \left(\sum_{i=1}^N s_i^d \frac{p_i^c}{p_i^d} \right)^{-1} = P_L^{dc}; \quad P_F^{cd} = \sqrt{P_L^{cd} P_P^{cd}} = \sqrt{P_L^{cd} / P_L^{dc}} \quad (2)$$

where s_i^c is the share of expenditure devoted to good i in country c . Note that, in ICP practice, these shares come from the National Accounts of each country, while the prices are collected by the ICP.

We need one more step to convert the bilateral indexes into multilateral indexes, which we discuss below. But that final step is less important than understanding the implications of (2), and the conceptual problems that are involved in choosing one particular index formula, such as the Fisher. When relative prices differ across countries, different index number formulas will give different answers, and if we cannot rely on identical tastes and cost-of-living indexes, we have a wide margin of choice, which can be thought of as a large margin of statistical uncertainty. One standard way of assessing the size of that margin is to look at the ratio of the Laspeyres price index to the Paasche price index, the ratio that is being “resolved” by using the geometric mean that gives the Fisher index. Table 1 shows the Paasche and Laspeyres index between the US and selected other countries, in the left panel, and between Nigeria and selected other countries, in the right panel. These are prices indexes for GDP excluding the balance of foreign trade calculated from the 2005 ICP data. Not surprisingly, Canada and Western European economies are closest to the US, and the Laspeyres indexes for prices in those countries relative to the US are only a few points greater than the Paasche indexes. Similarly, Nigeria is “close” to a number of its African neighbors, and more surprisingly, three countries in Eastern Europe, Latvia, Albania, and Estonia. But it is among the countries in the next panel that we see the problem. The US-based Laspeyres indexes for Tajikistan and Kyrgyzstan are 9.6 and 5.1 times the corresponding Paasche index; and although these two countries—especially Tajikistan—are outliers—other countries in

Africa and the CIS have ratios more than 2. With spreads this large, the choice of price index can make a very large difference in bilateral comparisons, and in some cases, we might reasonably doubt whether the data support any such calculation. We return to these issues below.

Equations (2) give us a set of bilateral superlative indexes which need to be made into multilateral indexes. If we denote by A the M by M matrix of logarithms of the Fisher indexes, we have guaranteed—by the superlative property—that A is skew-symmetric, and it has zeros along the diagonal. However, in general it will not be transitive, in the sense that, for all c , d and e ,

$$a_{ce} + a_{ed} = a_{cd} \quad (3)$$

which is what we need. (This property is sometimes referred to as “circularity,” see Diewert, 1987 for earlier references.) It is straightforward to show that transitivity holds if, and only if, there are M numbers, interpretable as the logarithms of the PPPs, such that, for all c and d ,

$$a_{cd} = b_d - b_c \quad (4)$$

There is no principled (backed by economic theory) way of enforcing (4). Corrado Gini (1924) suggested choosing b to fit the calculated A by minimization of least squares distance, a suggestion repeated later by O. Eltető and P. Köves (1964), and B. J. Szulc (1964), after whom this EKS method is (somewhat unjustly) named; GEKS is more appropriate and we use that term here. Given that we need to choose $b_1 = 0$ for the base country, the solution is readily shown to be, in terms of the original price indexes,

$$P_F^c = \left(\prod_{j=1}^M P_F^{1j} P_F^{jc} \right)^{\frac{1}{M}} \quad (5)$$

so that the GEKS-Fisher multilateral index, one for each country, is derived by taking the geometric mean over all of the possible M “indirect” Fisher indexes from the base country to the country in question.

Many PPPs, such as the Eurostat-OECD and World Bank PPPs, are calculated using some version of the GEKS method outlined above. In practice, it is surely impossible to do without the transitivity assumption; we cannot feasibly work with a *matrix* of price indexes. Yet transitivity comes at a serious price, specifically that the price index for any pair of countries depends on prices and budget shares in third countries, a violation of “the independence of irrelevant country” property. Indeed, Matthijs Van Veelen (2002) has shown that, given other mild conditions, transitivity and the irrelevance property are mutually inconsistent. As has been known at least since Fisher, price indexes cannot satisfy all of the properties that our price-based intuition suggests for them: price *indexes* are not *prices*. One possible source of comfort is an observation based on experience, but without theoretical foundation except in special cases, W. F. Alterman, Diewert, and Robert C. Feenstra (1999), which is that the matrix of bilateral superlative indexes, such as the Fisher indexes with which we began, are usually close to being transitive without further adjustment, so that the GEKS step has little effect on the calculations. By the same token, comparisons between pairs of countries using GEKS price indexes are not very sensitive to prices or budget shares in third countries. We also have a nice compromise between statistical practice and cost-of-living theory in the sense that the Fisher indexes are superlative indexes with a COLI interpretation if we are prepared to make the assumption of identical tastes, at least for some countries.

These indexes have (at least) one compensating drawback, which is responsible for the fact that they have not been provided in the Penn World Table until the latest version. The Penn World Table is a set of national accounts, with consumption, investment, GDP and so on, and these national accounts satisfy the standard national accounting identities, for example that consumption plus investment plus government plus exports minus imports is equal to GDP. When each component is converted to PPP by using a GEKS index for each, and expenditures converted to international currency by division, these identities no longer hold. Components of aggregates in international prices do not sum to their aggregates in international prices. For some purposes such as the World Bank's poverty work which uses only the PPP for household consumption, this is of no consequence. And Eurostat estimates appear to be widely used, at least within government agencies, without satisfying this requirement, though OECD also produces alternative, additive, estimates with a lag. For example, in Eurostat's 2005 benchmark data for France, "real" GDP at average OECD prices is 0.27 percent less than the sum of individual and collective consumption, gross fixed capital formation, the change in stocks and net exports, OECD (2009, Table 1.7). But economists studying the structure and dynamics of macro economies might find the violation of identities to be disconcerting, and the PWT is constructed along principles that preserve them.

The aggregation formula used by PWT was proposed by Roy C. Geary (1958), and is usually referred to as the Geary-Khamis (GK) method, Salem Khamis (1972). In the GK system, the prices in each country are compared with those of an imaginary composite country, itself constructed from averaging the countries in the system. The GK PPP index

for country c is computed as a Paasche index that compares domestic prices with “world” prices, which are the prices of the composite so that, for $c = 1, 2$,

$$P_{GK}^c = \frac{\sum_{n=1}^N p_n^c q_n^c}{\sum_{n=1}^N \pi_n q_n^c} \quad (6)$$

where π_n is the world price of good n , which is itself defined as the quantity weighted average of the prices of good n in each country, expressed in the global currency:

$$\pi_n = \sum_{c=1}^M \frac{p_n^c}{P_{GK}^c} \frac{q_n^c}{\sum_{d=1}^M q_n^d} \quad (7)$$

Equations (6) and (7) must be solved simultaneously, which can be done iteratively, or as shown by Diewert (1999) as the solution to an eigenvalue problem.

The advantage of this Geary-Khamis system of PPPs is that it preserves aggregation; because it uses a world price for each good, each item of GDP is re-priced at the world price, and added up to give re-priced subgroups or totals. The GK system also has a number of disadvantages that need to be balanced against this. Unlike the indexes underlying the EKS approach, it is not superlative. In consequence, if the two countries had the same homothetic tastes, the Geary-Khamis index would not be a second-order approximation to the “true” cost-of-living index. If this were the main concern, Geary-Khamis could be replaced by Neary’s (2004) GAIA system which is a consumer-theory consistent version of Geary-Khamis. If we do not want to assume identical tastes, nor use the cost-of-living framework that would be arguably appropriate if tastes were identical, these objections to Geary-Khamis are not decisive.

The central problem with GK (or with other methods that use a one price vector for all countries) is that it uses a single set of relative prices to value consumption or GDP in all countries, no matter how different are the actual relative prices. More specifically in the case of GK, the quantity weighting of prices in (7) means that the country with the larger physical volume of consumption of a good gets greater weight in the construction of the composite world prices; in this sense, GK gives a “plutocratic” set of international prices. If, for example, we used Geary-Khamis to compute a PPP for Bangladesh relative to the US, the world prices would be close to those of the US. In the Penn World Table as a whole, Daniel Nuxoll (1994) has argued that the composite world prices are those that would characterize a relatively rich country such as Italy or Hungary. Such prices are unlikely to be useful for comparing two countries whose relative prices are quite different from those in Italy or Hungary. Beyond that, the use of such plutocratic prices is likely to overstate the level of consumption in poor countries. For example, many services—haircuts, domestic service, restaurant meals—are cheap in poor countries because people are poor, because such services cannot be traded, and because labor is not free to move around the world. If we use (say) Italian prices to value (for example) Indian consumption, these components of consumption will be valued very highly, and will inflate the value of Indian consumption at international prices. This is called the Gershenkron effect after Alexander Gershenkron (1947), although once again Gini deserves priority; it is the overvaluation of one country’s consumption when evaluated at another country’s prices.

Doris M Iklé (1972) proposed an alternative to GK that shares its use of a single set of world prices, and that thereby preserves adding-up in national accounts, but replaces

the plutocratic weighting by aggregate quantities by a “democratic” weighting that gives each country equal weight. Iklé’s method, which was given little attention for many years, was recently clarified by Yuri Dikhanov (1997) and Bert Balk (1996), and Diewert (2008b) refers to it as the Iklé, Dikhanov, Balk (IDB) index. In the 2005 ICP, it was used to calculate the PPPs for the African region. Using Diewert’s formulation, the IDB index replaces the definition of the global price vector (7) by

$$\tilde{\pi}_n = \left[\frac{\sum_{c=1}^M \left(\frac{P_n^c}{P_{IDB}^c} \right)^{-1} S_n^c}{\sum_{c=1}^M S_n^c} \right]^{-1} \quad (8)$$

where P_{IDB}^c is the IDB PPP index for country c , which is calculated simultaneously with (8) using the Paasche formula (6) with $\tilde{\pi}_n$ replacing π_n . Equation (8) defines the global relative prices as budget-share weighted harmonic means of each country’s relative prices. Although the IDB index, like the GK index, uses the same relative prices for all countries, those relative prices are closer to those of poor countries than are the plutocratically-weighted GK prices; these prices are likely to be more relevant for comparing income levels in poor countries, and will reduce the size of the Gershenkron effect for the world as a whole.

If we compare the Paasche index in (6) with the superlative indexes presented earlier, the latter always averages weights from *both* countries, whereas the GK and IDB indexes use only *domestic* weights, and this is what generates the Gershenkron effect. Goods from rich countries are often rare and expensive in poor countries, if they exist at all, and goods which are rare in countries with no taste for them—alcohol in Muslim countries or English sausages in the US—can be very expensive when they are found at all. If alcohol

has a small share in Bangladesh, but a high relative price, weighting that price relative by Bangladeshi budget shares will understate Bangladeshi prices, while weighting it by OECD shares will greatly overstate them. The superlative indexes, which combine the weights, make a compromise that is arguably the best that can be done in the circumstances.

Compared with the superlative indexes, GK indexes will tend to understate PPPs in poor countries relative to rich ones, and thus tend to overstate their living standards. They make the world look too equal, and understate poverty in the poorest countries. For analyses of the world distribution of income or of world poverty, this would militate against using GK indices and in favor of GEKS or even IDB type methods. For work on growth or other macroeconomic questions, the additivity properties of GK or IDB may be more important, and the GK deficiencies may not be serious when the analysis is dominated by relatively similar rich countries. When such analyses involve poor countries in a substantial way, it should be borne in mind that the GK international prices that are used to value their goods and services are biased towards rich country prices, with the risks that this entails, for example in overvaluing cheap goods and services that have relatively little domestic value. For these, IDB is worth serious consideration.

Table 2 illustrates the various indexes for the same selected set of countries as in Table 1. Column 1 repeats the Laspeyres–Paasche spread as a reminder of where we would expect the different multilateral indexes to differ; the GEKS indexes are modified Fisher indexes, while the GK and IDB are modified Paasche indexes and will inherit at least some of their properties. We also show the bilateral Fisher index with the US as base, and then our own calculations of the GEKS, GK and IDB indexes using 128 basic

headings of GDP, excluding only the trade balance. The final column is the PPP index from the World Development Indicators 2008, which we shall discuss below.

There are several points to notice. First, the bilateral Fisher indexes and the GEKS indexes are usually not far apart, so that the GEKS modification to the bilateral indexes is not having much effect; there are exceptions such as Chad and Zimbabwe where the Laspeyres–Paasche spread is large, the data are weak, or both. If this limited difference holds more generally—as we suspect is the case—the GEKS indexes can be thought of much as we would think of Fisher indexes. The GK indexes are further away from both Fisher and GEKS indexes than the latter are from one another. Even so, the differences are small for similar countries where the Laspeyres–Paasche spread is small, but can become substantial where the spreads are large, or in other cases where the data are weak, such as Zimbabwe or Tajikistan. In the worst case—Tajikistan—the GK index is only 55 percent of the GEKS. The ratio of GEKS to GK is 1.056 in Nigeria, which is probably the worst of the large countries. The IDB index is larger than the GK index (countries are poorer relative to the US numeraire) for all countries in the world except Cyprus; the Gershenkron effect is consistently less for the IDB than the GK index.

Second, the GEKS indexes are consistently larger than the GK indexes for the middle group of largely poor countries. The Gershenkron effect is most apparent here, though we emphasize that many other factors are at work, most notably differences in consumption patterns that are not attributable to relative prices, as well as data errors. For the bottom group, there is no consistent inequality relationship between the GEKS and GK indexes; the contributions of these countries to world production are now large enough to remove any consistent Gershenkron effect for countries in the middle of the income distribution.

Third, for the second group of countries, the IDB price indexes are sometimes smaller and sometimes larger than the GEKS indexes while for the “other important country” group, the IDB indexes are always larger than the GEKS, so that, with democratic weighting, the Gershenkron effect has switched direction for the largest middle income countries.

Whether these differences are significant depends on the use to which the data are to be put, on which we shall have more to say below. One illustration comes from looking at world inequality, or at least the (dominant) between-country component of world inequality. We compute this from the local currency value of per capita GDP, and deflate by the three indexes. As predicted, inequality is smaller using the GK PPP; the population weighted gini coefficient for per capita GDP is 0.533 for EKS, compared with 0.527 using GK. With the IDB index, this Gini is 0.542, even larger than the EKS.

3. Operational issues with major implications for use of the data

3.1 Prices of items, prices of basic headings, and quality

The ICP collects and constructs the prices that go into the formulas, and although we try to avoid much of the complexity, some understanding of this process is necessary. While we emphasize operational issues, almost all raise theoretical and conceptual questions.

We begin with the regional organization of the ICP. Since 1980, the ICP has had a regional structure; one reason is logistical—local offices can better organize local data collection—and such a structure also allows the ICP to take advantage of data and expertise that already exist, for example in Eurostat who produce their own regional PPPs for other purposes. More substantively, PPP comparisons are almost certainly more

reliable between countries that have a similar economic structure, whether through tastes or conditions of production, so that regional systems of PPPs are probably more reliable than global systems, indeed the relative reliability of comparisons between close neighbors is clear from Table 1. In 2005, the ICP was decentralized into Africa, Asia-Pacific, Commonwealth of Independent States (CIS), South America, Western-Asia, and Eurostat-OECD. In the first stage of the ICP, each region collected prices using its own detailed list of goods and services. Each of these lists further disaggregates the common global list of 128 “basic headings” of GDP, “basic headings” being defined as the most disaggregated level at which there exist matching expenditure data from the national accounts.

The detailed prices for the goods and services below the basic heading level were then used to calculate a regional set of price indexes or parities for each of the basic headings, as well as a regional set of PPPs. Each region had its own numeraire country, sometimes a real country—Argentina in South America—and sometimes an imaginary composite country. At the end of this first stage, the ICP has a set of parities for each basic heading (sometimes also called PPPs, though they are commodity specific rather than country-wide PPPs), in which the parity for each good in Argentina (say) is 1, while for Chile, Brazil, and the other South American countries, there is a parity for each basic head which can be thought of as the basic-heading specific PPP exchange rate relative to Argentina. It is these parities that are taken to the formulas in Section 2, with expenditure data from the national accounts providing the weights, to yield an overall PPP for each country relative to Argentina. The same is done for each of the regions. Because the lists are not the same for each region, it is impossible to make direct comparisons between

countries in different regions, so that a separate method had to be developed to link the regions into a global system; this is the “ring” procedure, which is discussed in Section 3.2 below.

To illustrate from the final set of global calculations, “rice” is one basic heading in the consumption account. Some country parities for rice from the 2005 round are 4,304 Vietnamese dong per dollar, 0.65 British pounds per dollar, or 44.6 Kenyan shillings per dollar. If rice were the only component of consumption (or GDP), these would be the PPP exchange rates for those countries relative to the US; in fact, the actual consumption (GDP) PPPs for those countries are 5,920 (4,713) Viet Nam, 0.66 (0.65) UK, and 32.7 (29.5) for Kenya. Clearly, knowledge of the price of one good, or at least one group of goods, takes us some way, which is why the Economist’s Big Mac index is useful; of course, relative prices differ greatly from one country to another, which is why the Big Mac index is far from sufficient (or safe), and the ICP tries to do better by covering all the expenditures in GDP.

Below the basic heading level, there are no expenditure or quantity data, so the detailed prices have to somehow be aggregated up to prices for basic headings without weights, and the way in which this first-stage is handled turns out to have important consequences for the end result. As always, this two-stage process parallels the construction of domestic price indexes, such as the US CPI, where many of the same issues arise. The formulas for aggregating prices into basic heads differed somewhat from region to region, but looking at one of the methods is instructive. Suppose that in region r , for basic head i (these are the same goods and services in Section 2), we have $j = 1, \dots, N_i^r$ price quotes for goods within the basic head, e.g. six different kinds of rice,

long-grained, short-grained, brown, etc. In the absence of weights, there are combined into a geometric price index for the basic heading in country c

$$p_i^c = \left(\prod_{j=1}^{N_i^c} p_{ij}^c \right)^{1/N_i^c} \quad (9)$$

and the regional parity is given by (9) for c divided by (9) for the numeraire country in the region. (Note that, because of the structure of (9), these parities are both symmetric and transitive across countries.) In practice, except for the OECD which uses a somewhat different method, (9) is replaced by a “country product dummy” (CPD) regression of the form

$$\ln p_{ij}^c = \alpha_i^c + \beta_j^c + \varepsilon_{ij}^c \quad (10)$$

in which the logarithms of the price of each variety is projected on to a dummy for the country, α_i^c , and a dummy for the variety β_j^c . For each region, regression (10) is run for each basic head i , over data pooled over countries and varieties within the basic head for the regional list. The p_i^c in expression (9)—or divided by p_i^1 --is then replaced by $\exp(\alpha_i^c)$ in expression (10)—note that country 1 is the omitted country in (10). When all the varieties within the basic head are present in all countries, the two procedures are equivalent. But there are typically many missing values, when some countries are unable to price all items on the list, and (10) can deal with this, simply by including all available observations in the regression.

Note that (10) calculates parities for a basic head for two countries where the available items within the basic head do not overlap, provided there is at least one other country with both sets of items; for example, if there are two goods, one of which is

priced in country 1, p_1^1 , one in country 2, p_2^2 , and both in country 3, p_1^3 and p_2^3 , (10) calculates the parity for 2 relative to 1 as the product of p_1^3 / p_1^1 and p_2^2 / p_2^3 , which is the parity of 3 relative to one multiplied by the parity of 2 relative to 3, so that transitivity is used to fill in the blank. It could be argued that this is a “solution” to a problem that ought not to have one—it is essentially the rice and *teff* problem of the introduction—and it is certainly arbitrary in the sense that the parity between two countries depends entirely on information from third countries.

We can use these formulas to discuss the difficult issue of what happens when some goods are “representative” in some countries, but not in others. This is a milder version of the case where consumption patterns of different countries do not overlap. Consider two neighboring countries, one of which consumes wheat as its staple, and the other rice. In the wheat eating country, there is a substantial rice eating minority, and wheat and rice are easily available at the same price per kilo, p^1 (for country 1) say. In the rice eating country, very few people eat wheat, which is sold in specialty stores in the capital city at a price per kilo that is four times the price of rice. We thus have p^2 and $4p^2$ for the prices of wheat and rice in country 2. The regional list for the ICP for cereal contains both rice and wheat, so that when (9) is applied, the price of cereals is p^1 in country 1, and $2p^2$ in country 2, so that the cereal parity is $2p^2 / p^1$. Without the specialty shops, the cereal parity would be p^2 / p^1 , which is arguably the right answer. The underlying problem, of course, is the absence of expenditure weights below the basic head; if these were available, the unrepresentative consumption would get virtually no weight, and the calculation would give the right answer.

To deal with the issue of uncommon goods at high prices, Eurostat-OECD and the CIS developed procedures in which national statisticians judgmentally marked items as either representative or not representative of consumption in each country, and these ratings are used to downweight unrepresentative goods in formulas different from but with the same function as (9) and (10), see World Bank (2008, pp 157–8) for the details of Eurostat’s procedure. In the 2005 round of the ICP, it was planned to use representativity-weighting for all of the regions, for example by running weighted versions of (10), but the responses of countries were not encouraging; for example, all items priced in a country were often designated representative. As a result, no such corrections were made for regions outside the OECD and CIS.

One persistent criticism of ICP rounds prior to 2005 has been that the quality of items priced has not been strictly enough controlled, so that lower quality items in poor countries were often matched to higher quality items in rich countries, leading to an understatement of price levels in poor countries and to an overstatement of their output and income levels, World Bank (1993). This concern is more pervasive than just the worry that brain-surgery in Nairobi is unlikely to be identical to brain-surgery in Geneva. For many goods, the outlets sampled in poor countries may be closer to discount stores than to the typical outlet in the US or other rich countries. In consequence, successive rounds of the ICP have developed ever more precise descriptions of the goods to be priced, leading in 2005 to a formal set of descriptions known as “structured product descriptions” (SPDs). These SPDs represent a major improvement in the description of products compared with earlier rounds. They were designed to follow the United Nations Classification of Individual Consumption According to Purpose, known as COICOP,

which made it more probable that countries would use the coding for their own national purposes. Illustrative examples, here taken from the African regional specifications, are

(a) Nescafé classic: product presentation, tin or glass jar, 100 grams: type, 100% Robusta: variety, instant coffee, caffeine, not decaffeinated: brand, Nestlé-Nescafé classic

(b) Boubou (item within women's clothing): product specification, no package, 1 unit: fibre type, cotton 100%: production, small scale: type, boubou: sleeve length, sleeveless: fabric design, brocade: details/features, embroidery (c) light bulb: product presentation, carton, 1 piece: type, regular: power 40 watts: brand name, indicate brand.

Each region of the ICP has a list of goods, with SPDs, and there are typically many items within each of the common basic headings. Even within a region, not all countries can price all items, and some of the regions contain countries at very different levels of development, or with very different patterns of consumption and relative prices: for example, the OECD-Eurostat region contains both Japan and Mexico, Africa contains South Africa, Guinea Bissau, and Tanzania, and Asia Pacific Hong Kong, Taiwan, India, and Nepal. Given this heterogeneity, the use of SPDs and the more precise definition of goods is likely to increase the number of missing values in the CPD, as well as the risks of finding some high and unrepresentative prices as in the rice and wheat example above. If those non-representative prices are well-distributed across all of the countries in the region, they may not cause serious distortion, even in the absence of a representativity correction. The more difficult case, to which we will return in Section 3.2, is when we are directly comparing rich and poor countries, and the unrepresentative high-end prices are all in the poor countries. But this is mainly an issue for comparisons between regions, not within them, and there is no doubt that for the prices of items in the regions, ICP 2005 did

a much more thorough job of comparing like with like, of validating prices, and of involving country participation than in previous world ICP rounds. One caveat concerns the fact that, as the ICP pricing has evolved from its CPI beginnings, many more SPDs have been developed for goods than for services, even personal services that can be bought in shops, in part because quality definition is much harder. Measured at an average of PPP and local prices, more than half of the countries in the ICP have a share of expenditures on services in GDP of between 45 and 65 percent, and finding prices for some of these is much more difficult, an issue to which we return in Section 3.3

3.2 Linking the regions

One of the most difficult issues in the ICP is how to link the regions. Doing so requires some kind of direct comparisons between rich countries and poor countries, or more generally between countries whose structures of GDP and relative prices are very different. In the previous 1993 round, the linking was never satisfactorily completed, because different regions were linked in different years, because the bridging was done retrospectively, rather than as part of the original exercise, and because the lists of basic headings were not comparable across regions. The idea in that round was to work with “bridge” countries that appeared in more than one region, just as bridge periods are used to link time-series of prices on different bases. Another major issue was the sensitivity of the linking to the economic structure of the bridge country, so that in cases where there was more than one possible bridge country, the results were sensitive to which one was chosen, as well as to the level (basic heads or whole country) at which the linking was done. In the 2005 round, the bridging method was replaced by a ring of 18 countries,

chosen so that there were two or more in each region, and those countries priced a specially constructed common ring list of more than 1,000 consumption items; the ring prices were then used to link the regions. (Although it was not planned in advance, Russia was treated as a bridge; it was included in both the EU-OECD and CIS regions, and priced both lists. It was then used to link the two regions into a single EU-OECD-CIS region with a single set of basic-heading parities and country PPPs relative to one numeraire.)

The ring linking follows a procedure proposed by Diewert (2008a). For each basic head i in country c and region r , the ICP runs a CPD regression of the form

$$\ln \tilde{p}_{ij}^{cr} - \ln p_i^{cr} = \beta_{ri} + \gamma_{ij} + \varepsilon_{ij}^{cr} \quad (11)$$

where \tilde{p}_{ij}^{cr} is the ring price of good j within basic heading i in country c of region r . The prices are converted into the regional numeraire for the relevant basic heading using the regional basic heading parities p_i^{cr} , and their logarithms projected on to a set of regional and basic head dummies. The quantity $\exp(\beta_{ri})$, which will be unity in the numeraire region—the omitted region in (11)—is the overall regional parity for basic heading i . Matching aggregate regional expenditures are calculated from the expenditures for each basic head in each country converted to each region’s numeraire using the regional basic heading parities, so that

$$e_i^r = \sum_{c=1}^{C_r} e_i^{cr} / p_i^{cr} \quad (12)$$

where e_i^{cr} is the local currency expenditure (quantity times price) on basic head i in country c in region r , and the basic heading parities p_i^{cr} from the regional calculations are used to convert these into regional numeraire units. For example, if the region is South

America, (12) is total expenditure in South America on basic heading i in Argentinean prices. The regional aggregate basic heading expenditures e_i^r from (12), and the regional aggregate basic heading prices $\exp(\beta_{ri})$ are then slotted into the GEKS formula to give a set of global PPPs for each region. This regional PPP is 1 for the numeraire region, say OECD-Eurostat within which the US is numeraire, with the other numbers used to scale up the regional numeraire countries, e.g. Argentina, from 1 to its PPP relative to the US dollar, which comes directly from the ring GEKS. The global parities for each basic heading for each country are also calculated as the regional parities scaled by the regional basic heading parity, $\exp(\beta_{ri})$.

Diewert's procedure is only one of many possible ways of using the ring prices. For example, in the African region, the ring countries are Cameroon, Egypt, Kenya, Senegal, South Africa, and Zambia, and the ring prices can be used to calculate basic heading parities for this group of countries which are different from the basic heading parities calculated in the original regional exercise. In effect, (11) privileges the original parities, discarding the information from the ring prices, except to calculate the regional factors, $\exp(\beta_{ri})$, in this case for all of Africa. The main reason this was done is to respect the principle of "fixity," that the regional parities for each basic heading be incorporated into the ICP but not be disturbed by them. Starting in the 1980 round with Eurostat, whose freedom of action is restricted by EU regulation, all regions from 1985 onward made their participation in the ICP dependent on the condition that their own regional PPPs were respected in the final calculations. This is essentially a political, not a statistical constraint—for example, social funds within the EU are distributed on the basis of per

capita GDP at PPP—that must be respected by the World Bank in the official ICP calculations,

The global parities for the basic headings for each country, which are available to researchers for further analysis, have one aspect of fixity built into them through equation (11) and cannot be recalculated—for example by a single global calculation—without access to the underlying regional data, which are *not* available. Similarly, the forthcoming new versions of the PWT must use those parities for the basic headings, even though it is under no obligation to respect regional fixity in the future, and it has not done so in the past. But at the very least, not enforcing fixity within regions is one of the reasons why the PWT differs from other PPP data sets. One calculation is perhaps instructive. The 128 by 152 matrix of global parities by basic headings and countries can be used to recalculate a set of GEKS PPPs for all countries in one step, without respect to regional fixity, and the *broad* picture shown by these results does not differ much from the official ICP. However, for some countries the differences are not trivial, for example, the PPP for China is 6.6 percent lower (real GDP 6.6 percent higher) in a GEKS global calculation without fixity than in a GEKS two-stage calculation with fixity imposed. Further experimentation on alternatives to (11) using the ring data—which is also available to researchers—is clearly desirable.

In the limit, if every country were to demand fixity of its price level relative to all other countries, transitivity would be impossible, and we would not have an international system of accounts. Actual fixity is less severe, but it also places restrictions on the way that transitivity is imposed, and so will cause the final global system of PPPs to be further away from the matrix of pairwise superlative Fisher indexes that would be the basis for

an unrestricted calculation for all countries simultaneously. On the other hand, it is also clear that the Fisher indexes between two countries in different regions, between the US and Tajikistan, or between Nigeria and Japan, say, are worth less than those between the US and Canada, or Nigeria and South Africa, see again Table 1, and fixity recognizes this fact, albeit in a crude way. The PPPs from the ICP, which incorporate the fixity constraints, are listed in the final column of Table 2. For several of the countries, these estimates are outliers relative to the other indexes shown. This is true, not only where it is to be expected, in the countries with very large Laspeyres-Paasche spreads, but also in some of the important countries in the bottom panel. For example, the PPP for China in the ICP is more than ten percent higher than the EKS or bilateral Fisher index, while that for India is 7.5 percent higher. These results are somewhat puzzling given the EKS basis of the Asian numbers, more so than the also large difference for Nigeria, where the ICP does not use the EKS method. We suspect that the interaction of the productivity adjustment (discussed in Section 3.4 below) and the regional aggregation may be largely responsible for the differences. It is perhaps also worth noting that the use of the PPPs from the ICP also has a mild positive effect on measures of inequality between countries; the population weighted gini coefficient of GDP per head, which was 0.533 and 0.527 for the EKS and GK becomes 0.580 using the ICP official data.

Perhaps the most difficult aspect of linking the regions is not the fixity constraint, but the fact that the ring must price identical items in all of the countries. In addition to the six African countries already listed, the ring included the UK, Japan, Slovenia, and Estonia from the Eurostat-CIS region, Hong Kong, Malaysia, Philippines, and Sri Lanka from Asia-Pacific, Brazil and Chile from South America, and Jordan and Oman from

Western Asia. It is clearly a considerable challenge to develop a list of more than a thousand items to be priced in all countries in such a group. It is also here that we face the sharpest tradeoff between, on the one hand, matching quality by pricing identical items using the detailed SPDs and, on the other, risking the high price unrepresentative good problem by pricing rich country goods in poor countries. Note too that (11), like (9) and (10), aggregates item prices up to basic heads, so that there are no expenditure weights, nor is there any representativity correction to offset prices collected in the occasional high-end shops that are added to the list of usual CPI outlets for ICP purposes.

Again, to illustrate, ring items that were *successfully* priced in Cameroon included frozen shrimp (Fish basic heading: 90-120 shrimp per kilo, pre-packed, peeled), Bordeaux red wine (Wine basic heading: Bordeaux supérieure, with state certification of origin and quality, alcohol content 11-13%, vintage 2003 or 2004, with region and wine farmer listed), Heineken (Beer basic heading: 0.33 to 0.5 liter bottle), and a frontloading washing machine (Major household appliances whether electric or not basic heading: capacity 6 kg, energy efficiency class A, Electronic program selection, free selectable temperature, spin speed up to 1200 rpm, medium cluster well-known brand such as Whirlpool.) If the expenditure share of wine in Cameroon is small, the choice of items within it is not very important, though recall that the GEKS calculations begin from bilateral Fisher indexes that use weights from both countries, so that the price of wine in Cameroon will attract some of the much larger French weight on wine. However, if almost everyone in Cameroon drinks local beer (and domestic lager beer is an item in the ring list), the calculated parity for beer will give the same weight to Heineken as to the domestic beer, which will help overstate the price level in Cameroon, and help understate

the size of its economy. (Note that working only with domestic beer does not resolve the problem if domestic beer in Cameroon is of lower quality than domestic beer in Germany or the Czech Republic.) The same considerations apply to the basic heading for appliances. It is perhaps not surprising therefore that, when country parities are calculated using only the ring data, Cameroon's price level—the ratio of its PPP to market exchange rate—was higher than that of the much richer Hong Kong. This may be more accurate as an estimate of the relative costs of a Hong Kong businessman posted to Cameroon than as an estimate of the relative costs of living in the two countries.

Our intent here is to highlight the inherent difficulty of making price index comparisons, not to criticize the ICP procedures in 2005, which were vastly better than those in earlier rounds. Indeed, the clarity of these pricing problems in the 2005 round is itself a compliment to the quality of the data collected, and the range of countries covered. In this respect, the regional structure and the fixity requirement might actually be helpful, because (11) makes the minimum use of the ring prices, using them only to link the regions, so that the ring prices enter into the final calculations only through the estimates of $\exp(\beta_i^r)$. Even so, there is a conflict between the search for more precise quality matching, on the one hand, and representativity on the other, and this is not resolvable without expenditure weights below the basic heading level or, more fundamentally, a better developed theoretical idea of what we mean by a comparison of the cost-of-living between Cameroon and Hong Kong, or Oman and Japan. In the meantime, we judge it likely that the ring comparisons in 2005 have overstated the price level in the poorer regions, and have thus exaggerated the inequality of PPP per capita incomes between rich and poor countries, certainly compared with previous rounds, and

probably compared with better measures. But this judgment is far from solidly established on the current evidence, see also Deaton (2010).

3.3 China, India and other large countries

Until 2005, China had never participated in a full ICP round, though China did participate in a limited comparison in the 1993 ICP between Beijing and Hong Kong under the regional coordination of the Economic and Social Commission for Asia and the Pacific, ESCAP (1999). Consequently in all versions of the PWT to date, the Chinese numbers were estimated using partial information and shortcut methods. India, which also participated in 2005, had not previously done so since 1985, and its prices since then have been estimated by a mixture of updating and regression methods of the kind described in Section 4 below. The estimates for both India and China, whether in the PWT, or other PPP databases, are therefore relatively uncertain compared with more recently benchmarked countries. The Chinese and Indian data from the 2005 round, which will be incorporated into version 7.0 of the Penn World Table, are therefore of more than usual interest. Discussion of China and India also brings up an important general issue, which is differences in prices across space, both between different cities, and between urban and rural sectors within a country. In principle, prices for the ICP are national average prices, but in practice, rural prices are not always collected, something that is a more serious issue in large economies than in small ones.

We start with China, where the 2005 price levels are much higher than previously projected, so that there is an immediate question whether prices were overstated through some combination of choice of brand, outlet or location, though it is also possible that the

data underlying the previous price projections were understated. Price collection by China in 2005 took place in 11 cities and in their immediately surrounding areas which are mostly urban, but with some rural characteristics. Apart from these, no rural prices were collected. Reported price differences between the 11 cities or between the cities and their surrounding rural areas were not reported, but they were apparently not large. Because the Chinese expenditure data refer to the whole country, the Asian Development Bank (ADB), which was the relevant regional authority, corrected the collected Chinese prices to put them on an all-China basis. The ADB convened an Expert Group (Asian Development Bank, 2006) who decided to make no adjustment of the 11 city prices, though it did weight the 11 city prices according to the expenditures of the provinces to which they assigned on the basis of clustering of census characteristics. This adjustment takes no account of possible differences between urban and rural prices.

That urban prices are substantially higher than rural prices has been documented by Loren Brandt and Carsten A. Holz (2006, Table 7), Shaohua Chen and Martin Ravallion (2008) and by Cathy Honge Gong and Xing Meng (2008). Chen and Ravallion also note that the Chinese National Bureau of Statistics chose the 11 cities because they were most likely to have outlets carrying the types of products and brands in the ICP specifications, and those prices are likely to be unrepresentatively high. The provinces in which the 11 cities were located had rural and urban incomes substantially above the all China average, suggesting that there would be regional price differences. In addition, the counties in which the 11 cities are included provided poverty claims to their residents that were 50 percent higher than the national average. Taking the various estimates together, we think a conservative adjustment would be to treat the urban prices as collected in the 2005 ICP

as 20 percent higher than national prices. Taking into account the distribution of consumption between urban and rural areas, this adjustment would raise estimates of Chinese GDP in 2005 by about 10 percent.

Do similar spatial differences exist in other large countries? Bettina 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 percent of the US average in 2003. From more than a million collected prices, Aten is able to obtain about 25,000 annual average price observations for 256 entry-level items collected by the BLS and uses these to estimate price level differences over all of consumption. This is a rich data set that has now been updated annually up to 2007 with similar findings, so that we can be fairly certain that the range across US urban areas is around 60 percent, suggesting that for China, where spatial price differences are almost certainly larger, a 20 percent downward adjustment is conservative. 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, nor in the expenditure surveys that underlie many of the studies of Chinese price differences.

There is a clear problem for the ICP in comparing large versus small countries. If all countries had the same ratio of urban to rural prices, the PPPs based on urban samples of prices would be the same as those for national prices. For Belize, Bermuda, Hong Kong, Luxembourg or Singapore, the frame of outlets for the CPI covers the country and provides a good basis for the ICP. But this is much less true for large countries, at least

for those that have a substantial share of their population and transactions in rural areas. Not only are large parts of these countries not covered in the CPI, but the sample of outlets is typically not well suited to the ICP's lists of items—the problem of comparable goods once again. In 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 typically collected urban prices. Even in India, the usual outlets for collection of rural prices would not typically include outlets in smaller regional centers where some of the ICP items are more likely to be available.

In proposing that Chinese prices should be adjusted it is recognized that this might not be appropriate for comparisons with other Asian countries that primarily rely on city prices, and it is these comparisons that most concern the ADB, for example. However, comparisons of China with the G-20 countries are also an important dimension of the global ICP and for this purpose a downward adjustment of Chinese national prices seems appropriate. The general problem of large countries providing estimates of national average prices remains a problem to which resources will be put in the 2011 ICP round

3.4 Comparison resistant items: housing and government services

The ICP uses the term “comparison resistant” to refer to goods and services for which it is difficult or impossible to observe market prices that can be compared across countries. Examples are housing rental, government services, health, and education. Measurement of these is problematic for the national accounts within countries, but becomes still more hazardous in international comparisons. We focus on services provided by government

and services from the housing stock, but will have something to say about health and education too. Many macroeconomists using these data may be interested only in aggregates such as national income or consumption, and not in these items for their own sake. But the measurement problems for these items are large enough to have major effects on the larger aggregates; indeed we suspect that the largest single factor responsible for the decline in the relative size of the Indian and Chinese economies is a change in the treatment of government services.

How does one compare the output of civil servants and health and education workers across countries? Within countries over time, national accounts deflate nominal salaries to a quantity basis by making assumptions about changes in the productivity of the relevant workers. Similarly, in the ICP, assumptions about differential productivity across space fill in for the missing prices. In past ICP rounds, volumes were derived by dividing compensation by a PPP derived from a detailed comparison of salaries for occupations typical of government staff expenditures. Such an assumption implies that productivity in the provision of these services is identical across countries in a given occupation, which is unlikely given very different amounts of accompanying capital per worker across countries. Further, there is little 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, so that the consequences of the equal-productivity assumption loomed much larger. In Asia for example, salaries for government health workers measured at market exchange rates are 120 times larger in Hong Kong than in Laos; similar differences exist

between Yemen and Kuwait in the Western Asia region. If we assume that productivity is the same, per capita volumes of these comparative-resistant services in Yemen or Laos greatly exceed those of its richer neighbors, an improbable finding. The situation in Tajikistan is similar and helps account for its large Paasche-Laspeyres spread in Table 1. Such adjustments have been considered earlier by the OECD and the ICP, but the 2005 comparisons in the Asia-Pacific, Africa, and Western Asia regions are the first actual cases where the equal productivity assumption has been significantly modified. Productivity differences in OECD-Eurostat, CIS, and South America were not dealt with, either because they were thought to be too small to warrant it or because there was no agreement on how the correction should be made.

Irving B. Kravis, Heston and Summers (1982, Chapter 5) discussed comparison resistant services for 34 benchmark countries in 1975. A comparison was made between the PPPs for priced services and non-priced services, based on input prices. Priced services were higher than non-priced services in low income countries with the relationship reversed in higher income countries. The OECD considered using this relationship to estimate PPPs for non-priced services, but eventually decided against it partly because many countries did not price enough market services. The OECD also considered making productivity adjustments on the basis of reference PPPs but could not agree on a procedure. In the 2011 ICP the OECD will make estimates of educational output based on students at various educational levels with some quality adjustment based on standardized tests. For rounds beyond 2011 they have begun research on the health sector.

In the 2005 ICP, Asia, West Asia and Africa made productivity adjustments based on estimates of capital per worker in each economy as a whole. In Asia, for example, this has meant that the volume of GDP of China and India relative to Hong Kong or Singapore is lower than in previous ICP rounds. For example, government services per capita in China are 54.9 percent of Hong Kong without the productivity adjustment and 24.2 percent of Hong Kong without it. There is therefore a comparability issue across regions in 2005 because Europe, the OECD, the CIS, and South America made no 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 region and the ICP as a whole, Asian Development Bank (2007) and World Bank (2008a).

The productivity adjustment is clearly in the right direction relative to earlier benchmarks which attributed too large a volume of such services to poorer countries and biased upwards 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 likely significant error associated with the actual adjustments, even for countries within the same region. Further the 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 Western Asia would have been substantially lowered if the 2005 procedure had

been adopted in those earlier years. The earlier procedures have thus artificially compressed the distribution of income between countries.

The large decline in the size of the Chinese economy in 2005 compared with earlier benchmarks is in part attributable to the way that the productivity adjustment interacts with the requirement that regional parities be respected. The productivity correction in Asia changes China relative to Hong Kong or Singapore, but this correction can only have limited effect given regional fixity within the Asia Pacific region. If we remove fixity and calculate a global GEKS for all 146 countries, we introduce direct comparisons between compensation in collective government, education and health in China and in countries such as the US. A one-step GEKS calculation reduces the price level in China by 12 percent compared with the official ICP, about half of which is attributable to fixity, and the rest to other features of the ICP calculation. Our adjustment for rural prices discussed in Section 3.3 above reduces the Chinese price level by another 12 percent, so that the two adjustments together account for about half of the “shrinkage” in the size of China relative to estimates prior to the 2005 ICP. Of the remaining half, some should be attributed to the productivity adjustment itself (rather than its interaction with fixity), and some to weaker data prior to the 2005 round. Of course the productivity adjustment is also subject to a good deal of guesswork, so that we are not exactly on firm ground in making these comparisons.

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 (excluding Korea and Japan) was reduced compared to the OECD countries (here including Korea and Japan) as a consequence of the productivity adjustment. This would

also affect the comparisons of Asian countries with countries like Brazil, Mexico and many eastern European countries where the productivity adjustment was not carried out. This is not an argument against a productivity adjustment, though the actual implementation was perhaps insufficiently tailored to country specifics. But a knowledge of the actual adjustment helps us better understand why the position of China and India has changed so much in the recent round. And it certainly points to the need to gain agreement on a standard method of treating un-priced services for future rounds.

Another important comparison-resistant component of GDP is rental of housing, including both actual rents and the imputed rents of owner-occupiers. The share of housing rent in GDP is about 10 percent in the US, 9 percent in the UK, but only 4.7 percent in India, 2.2 percent in Nigeria, and an incredible 0.5 percent in Ghana; some of these differences are likely to reflect difficulties in measuring rent in the national accounts, or failure to make an imputation for owner-occupier rents.

Prior to 2005, ICP comparisons were based on surveys of rents, which allowed market rent comparisons for various size and amenity groups of housing and, assuming rental equivalence, for owner occupied housing. The EU and OECD countries used a similar survey 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 for such countries was to make direct comparisons of quality-adjusted volumes of housing, the “quantity” as opposed to the “survey” approach, and to find a link member country or countries, initially Austria, that would both survey rents and provide quantity information on their housing stock. Of course, other countries

had faced similar problems earlier, including those with rural housing stocks that are rarely if ever rented and had adopted a range of methods.

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 measurement was effectively abandoned in favor of the assumption that the per capita volume of housing services for each country was the same proportion of the world average volume of housing services as was 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 regions. Another is the bizarre consequence in countries, like Ghana, whose national accounts show little expenditure on rents. In these cases, the PPP for the rental category is calculated by dividing a very small number by a relatively large one, so that the parity for this basic heading is wildly out of line with the overall PPP. For Ghana, the parity for the rental heading is 178, less than 5 percent of the overall PPP of 3721. This is an extreme case, but there are others that are almost as extreme; Malawi also has a ratio of 5 percent, and Chad a ratio of 6.9 percent. Although the local (Ghana, Chad, Malawi) expenditure weights attached to these prices are also very small, the same is not true of the countries with which they are compared in calculating the bilateral Fisher indexes that go into the PPPs. For the three countries listed, their overall PPPs are reduced by close to ten percent comparing PPPs with and without the rental category, and the size of their estimated incomes inflated by the same

amount. Perhaps this is not very large given the overall quality of African GDP data, but it is worth bearing in mind.

We shall not discuss health and education here, but there is an associated issue that might not be familiar to all economists. The most recent versions of the United Nations' System of National Accounts (SNA), adopt a definition of household consumption, referred to as "actual consumption" which includes services provided for households by government and non-governmental organizations, including health and education. From a national accounts perspective, this makes a certain amount of sense, because it prevents the size of consumption depending on the extent to which health and education are publically or privately provided. However, there are many countries around the world where government-provided health and education is inefficient, sometimes involving mass absenteeism by teachers and health workers, Nazmul Chaudhury, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan and Halsey Rogers (2006), so that such "actual" consumption is anything but actual. To count the salaries of AWOL government employees as "actual" benefits to consumers adds statistical insult to original injury. The more traditional concept of household consumption, is sometimes presented—though sometimes hard to find—and is typically labeled "individual consumption expenditures by households" or some other term not containing the word "actual."

4. Filling in over space and over time

Since the ICP was begun in 1968, there has been an interest in covering non-benchmark countries, and PWT was launched to serve that interest. In addition the international agencies wish to cover as many of their member countries as possible. A brief discussion

of filling in non-benchmark estimates is given in section 4.1. This involves both spatial and temporal extrapolation. Section 4.2 is about the extent to which successive revisions of the PWT are compatible, and reports important lessons from recent research. In these subsections, we give most attention to the PWT, since it has been most heavily used in economic research, and because it is the only one of the databases that provides a long time series of disaggregated national accounts for the world as a whole. (Similar data for Europe and the OECD, stretching back to 1980, and including some higher frequency data—quarterly and monthly—come from the Eurostat-OECD PPP program. The World Bank’s World Development Indicators and the IMF World Economic Outlook also provide PPP related numbers for member countries for GDP and some of its components.)

4.1 Non-benchmark countries and updating over time

When non-benchmark estimates were launched in the mid 1970s the number of benchmark countries was only 16. In 2005 there are 146 benchmark countries, so the number of non-benchmark countries is about 40, many of which are in the Caribbean which was the only significant country grouping that did not participate. Different databases use different imputation procedures for the countries without data, although all rely on the strong positive relationship between national income and the price level of GDP, defined as the ratio of the PPP to the market exchange rate. Poorer countries are relatively richer in PPP terms, proximately because non-tradable goods are cheap where wages are cheap, and more fundamentally, because of the Balassa-Samuelson conjecture, that rich countries are relatively more productive in the traded-goods sector. Figure 1

shows this relationship for the 2005 round, plotting the logarithm of the price level of GDP, with the US as 0, against the logarithm of GDP per capita expressed in market exchange rates; the heteroskedasticity in the Figure may reflect data quality as much as failure of Balassa-Samuelson at low incomes. Each country is plotted with a circle whose diameter is proportional to population size; the gross outlier here is Zimbabwe. The PWT uses the log of the price level of domestic absorption as the dependent variable and adds, in addition to the logarithm of GDP per capita, direct (although admittedly imperfect) information on prices taken from foreign-posting cost of living adjustment indexes from the International Civil Service Commission (ICSC), as well as from the US and Canadian foreign offices, and an openness to trade variable that captures involvement in international financial flows, and regional dummies for Africa and the OECD. The World Bank, subsequent to the 2005 round, improved their approach, and these improved results are now incorporated into the WDI, Changqing Sun and Eric Swanson (2009).

The filling-in of non-benchmarked countries is today a much smaller problem than it has been in the past, and should be an even smaller issue in the 2011 round whose coverage will be larger still. However, even now there is a six year gap between the most recent and next ICP, and the gaps in the past have been larger. Years between benchmarks are filled in by extrapolation of country parities. The World Bank's calculations typically update the PPP exchange rates in the simplest way possible, using the domestic rates of inflation relative to the United States. Hence, if the PPP of country c at benchmark t is x units of local currency per dollar, the consumption PPP in year $t + k$ is given by

$$x_{t+k} = x_t \cdot (P_{t+k}^c / P_t^c) / (P_{t+k}^1 / P_t^1) \quad (13)$$

where P_t^c is the domestic price level in country c at time t , and country 1 is the numeraire country, in this case the US. The domestic price level will be the price deflator of GDP if we are updating the PPP for GDP, or the CPI if we are updating the PPP for consumption. In general, (13) will not match the new benchmark for many reasons, including the facts that the items priced in domestic price indexes do not match the international lists in the ICP, that local and ICP prices often have different geographical coverage, and that (13) is unilateral (or bilateral relative to the US) not multilateral, as in the benchmark.

In the PWT, updating according to (13) is carried out separately for consumption, investment, and government expenditure. These three PPPs are then combined their corresponding current price expenditures and a new (multilateral) Geary-Khamis aggregation is carried out. The net foreign balance in current prices is converted at the PPP for domestic absorption and added to domestic absorption to obtain GDP.

The PWT provides two distinct constant price (say 2000—this “reference” year need not be an ICP benchmark year) GDP series, a fixed weight index, and a chain weighted index which differ in the way that they are updated. The fixed weight index uses the share of C , I , and G in the reference years as the weights applied to national constant price growth rates for each of these expenditure aggregates; this is equivalent to growing each component of GDP at its local real growth rate, and adding. This fixed-weight procedure becomes less attractive the older the benchmark, which is the reason for the chain weighted index. This applies the current price weights of the year t as described in the previous paragraph to the growth of C , I and G between t and $t + 1$. This provides an estimate of the growth rate of domestic absorption between t and $t + 1$ which can then be

applied to domestic absorption in year t in constant (2000) prices. These procedures, together with (alternative) treatments of the net foreign balance, are discussed in more detail in Summers and Heston (1991, 343–44), and also in Simon Johnson, William Larson, Chris Papageorgiou, and Arvind Subramanian (2009). It should be noted that neither of the methods used in PWT for handling the foreign balance are fully satisfactory and indeed trade is an important area for future improvement in both the ICP and PWT.

A final issue worth noting is the treatment of the formerly communist countries of Eastern Europe and the former Soviet Union. Until events of 1990-91, PWT followed the prevailing practice of making estimates for “market economies”, and treating “non market economies” separately. Several of the latter, Hungary, Poland, Romania and Yugoslavia had participated in the ICP, but their constant price national accounts were not in a form that could be moved to non-benchmark years. The European non market economies and occasionally Cuba, though never China, had been carrying out purchasing power comparisons since the 1960s under the Council of Mutual Economic Assistance (aka COMECON) making binary comparisons with the Soviet Union. These studies were not officially published until the last comparison for 1990, but they appeared in journals and were the basis of the early treatment in PWT.

The integration of estimates of these countries into PWT began with PWT 5.6. However, there were major obstacles for those COMECON countries that had not previously recast their national accounts from the Material Products System into the SNA form. Further, price indexes over time for many of these countries were particularly difficult because the movement from administered to market prices posed a major data collection problem because there did not exist any framework for collecting prices in the

field. Hungary and Poland were well equipped for the transition, but to varying degrees the quality of data is uneven; those former Soviet Union (FSU) and Eastern European countries now members or associates of the EU have the most reliable data for ICP and PWT purposes. An exception would be those areas affected by the break-up of Yugoslavia. The Commonwealth of Independent States (CIS) and its statistical office CISSTAT (the former COMECON secretariat in Moscow) now coordinates PPP and other studies for the nine member countries. Russia, which also participates directly in OECD PPP studies, has a strong statistical apparatus, while the other countries are quite mixed as is illustrated by the case of Tajikistan in Tables 1 and 2. Further the constant price national accounts series for the COMECON countries are all much more reliable after the mid to late 1990s than earlier, with Hungary and Poland as (positive) exceptions.

4.2 Is PWT consistent across versions?

In his *Principles of Economics*, Marshall began with a variation of the Latin proverb, *natura non facit saltum*, there are no jumps in nature. Reviews of studies based on different versions of PWT raise questions about whether some of the jumps in PWT are real or rather in the nature of PWT's construction. 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. As we have seen from the discussion of the 2005 round of the ICP, substantial methodological changes continue to be made. National accounts are subject to revisions, and changing reference years of PWT also introduces elements of non-comparability between different versions of PWT. Users have been advised of these changes with each

update of PWT, but it is not always clear that the advice has been heeded. Figure 2 provides some evidence from a comparison of versions 6.1 and 6.2 of the PWT. These show annual growth rates of real chained per capita GDP, computed over ten year periods, starting in 1960, 1970, 1980, and 1990, with 6.1 on the horizontal axis, and 6.2 on the vertical axis. Although most countries cluster close to the 45-degree line, there are a substantial number of revisions, particularly in the decade beginning in 1970, and particularly—but by no means exclusively—for African countries. As we might expect, revisions are greater for the poorer countries. The decade beginning 1950, and not shown here, had very little revision between the two rounds.

Work by Johnson, Larson, Papageorgiou and Subramanian (2009, hereafter JLPS) sheds some light on the consequences of these revisions. One issue is differences in growth rates between PWT versions. Annual growth rates of GDP based on the chain series were compared by JLPS from PWT 6.1 and 6.2 for 40, 10 and annual intervals. Why would they differ? The discussion above suggests 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. Quoting JLPS (2009, p.7), “One basic aspect of data variability between PWT revisions—namely that it increases when the data are at higher frequency—becomes evident when we compute the growth rate over 1-, 10-, and 29--year periods.” Differences are lower for high-income countries, and larger for both low and middle-income countries, which is parallel with the grades assigned to country data quality in PWT.

Do these growth differences between versions of PWT make a difference? JLPS undertook an extensive literature search and replicated 13 published papers using

alternative versions to those used by the authors from PWT 5.6, PWT 6.1 and PWT 6.2.

The guidelines that emerge are:

- 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.
- SAFE: Use of annual growth rates for the A and B grade countries is safe.
- NOT SAFE: It is not safe to use annual growth rates from different versions of PWT for non-OECD countries.
- NOT SAFE: Different versions of PWT are not robust for dynamic analysis at annual frequencies.

A recent paper by Antonio Ciccone and Marek Jarocinski (2009) (CJ) identifies another type of analysis for which revisions in the PWT are a serious problem. These are attempts based on automatic model selection procedures to have the data choose which variables should appear in growth regressions. CJ apply these procedures to version 6.2 and 6.1 and obtain very different variables; in one of their experiments, the two data sets disagree on 13 out of 23 growth determinants, some of which are widely used in the literature. When they restrict the range of variables over which selection is possible, the results are a good deal more robust, if not perfectly so. These experiments may tell us as much or more about the failings of statistical model selection than about the failings of the PWT. Such procedures have are sensitive to small changes in the data that makes them unsuited for use with the PWT, none of which implies that the data are not well-suited to answering better-structured or more theoretically informed questions.

This brings us to the problem of integrating the new view of the world economy in the 2005 ICP with the older view in the WDI and PWT. The World Bank will simply begin anew with ICP 2005 and not consider earlier ICP rounds in future WDIs and other publications. When ICP 2011 is completed there will be a problem of reconciling the two rounds but until then, WDI extrapolations will be done at the GDP level. As far as PWT is concerned, the plan is to present a reasonable view of the world economy in 2005 and to move that backward and forward in time. The work of JLPS suggests that PWT should follow standard national statistical practice and provide a consistent set of accounts on a 2005 base in PWT 7.0, with full incorporation of the 2005 ICP. The JLPS research also suggests that more alternatives need to be considered for updating and backdating the PWT numbers through time. They also suggest that more consideration be given to earlier rounds of the ICP going back perhaps as far 1970; they argue that each round of ICP captures the structure of quantities and prices in that year across a spectrum of world countries. Finding a way to integrate these unique data sets across benchmarks, as will also need to be done when the 2011 round is complete, would be make a major contribution to our understanding of the growth process. It is easier to agree with this conclusion than to decide exactly how to implement it.

For the reasons already discussed, the results of ICP 2005 will not be woven into PWT 7.0 without adjustment. For example, the fixity restrictions will not be respected, and some preliminary analysis suggests that this will make a marked difference to some of the PPPs, see again Table 2. Other possible adjustments include modifications for the special character of Chinese prices, the lack of comparability of non-priced goods and services across the regions, and presentation of alternative aggregation methods. 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.

5. Conclusions

This summary of PPPs in both the ICP and the Penn World Table has covered only a fraction of the issues that go into the construction of these data. An example of an important omission is the unsatisfactory treatment of exports and imports, an issue that is unresolved, in the ICP, the PWT or other data, see Feenstra, Heston, Marcel P. Timmer and Haiyen Deng (2009) Experience suggests that it is hard to know in advance which features of the data are likely to be decisive for which purpose, or which particular detail will be responsible for some new or potentially interesting finding. Perhaps the overriding message is to exercise caution, particularly with comparisons between countries whose economies are very different, and particularly with the national accounts data provided by countries whose statistical capacity is weak. On the former, there are deep conceptual difficulties that cannot be resolved by collecting better data. On the latter, it must always be remembered that the international accounts are no better than the national accounts of the participating countries. The quality ratings in the PWT contain useful information that should be more heavily used.

There are also some specific health warnings that are worth emphasizing. One is about index numbers, and the general point that price indexes are not prices. Although most economists know that different price indexes give different answers, the comfort that comes from thinking that it matters little in practice is a strictly domestic comfort

that does not always travel well. Differences between Paasche and Laspeyres indexes are sometimes very large in the ICP, and these extend to differences between EKS and GK aggregations that are used by different agencies, see again Tables 1 and 2. Second, given the regional structure of the ICP, it is always worth taking into account the possibility that the regions are not comparable in some important respect. Regional dummies are often included in growth regressions for substantive reasons—or as an admission of ignorance—but there are also statistical reasons for including them, or even, when possible, treating the regions separately. Third, there are particular reasons for caution in using the data for countries of the CIS and the former Soviet Union. Fourth, some important components of GDP, including government services, health care, education, construction, and the rental of housing, are extremely difficult to compare across countries, and are often handled by sensible, but more or less arbitrary assumptions. Not only is it dangerous to rely on the benchmark estimates for these items, but their treatment can affect overall PPPs between countries, or even regions.

Researchers have a wide range of data sources. The World Bank's WDI contain time-series for GDP in current and constant international dollars, but not for the other components of the national accounts. Eurostat provides PPP accounts for Europe and the OECD, with GK versions following the EKS data with a one year lag. Time series data go back to 1980 in some cases, and there are some data at higher than annual frequency. For researchers who do not need data from Asia, Africa, or Latin America, these are worth serious consideration. The PWT, on which we have focused, is essentially the only option for long time series containing the main variables of the national accounts.

We end with a warning. Successive editions of the ICP have tended to revise upward the PPPs of poor countries relative to those of rich, even in the same year. Closer comparability of the goods and services to be priced is one of the reasons; most recently, more appropriate assumptions about the productivity of government workers has had the same effect. In particular, successive revisions have tended to make India and China poorer than the previous revision, though by no means at the same rate. For several years, in spite of China's much more rapid growth, the ratio of Chinese to Indian per capita GDP did not rise by as much as would seem warranted, leading to suspicions that the government of China wished simultaneously to exaggerate its growth rate and to understate its level of per capita GDP, see T. N. Srinivasan (1994, p.10). The recent (apparent) shrinkage of both India and China in the 2005 ICP has provoked similar concerns, most notably from Surjit Bhalla (2008), a long-time critic of the World Bank. Using the numbers here, a version of Bhalla's argument is as follows. The 2005 estimate of Chinese per capita GDP at 2005 international dollars is \$4,091. According to the PWT6.2, which is based on the Chinese official data, China grew at 5.52 percent a year from 1952 to 2004; at this rate, GDP per capita in 1952 would have been \$279 in 2005 international dollars, or \$153 at 1985 international dollars converted using the US CPI. Pritchett (1997) has persuasively argued that approximately \$250 in 1985 international dollars is the minimum level of per capita GDP that is required to sustain a population, or that has ever been observed for more than a short period. If so, it is simply not possible that *both* the current PPP estimate of Chinese GDP *and* the official growth rates of the economy can be correct. On the latter, Maddison (2007) estimates Chinese GDP growth since 1952 at "only" 4.4 percent a year, but this still leads to \$229 in 1985 international

dollars, still below Pritchett's cutoff. Reducing the PPP by twenty percent or so, as suggested in Section 3.4 above, would bring this number into a somewhat more plausible range. More broadly, the point remains that many of these numbers have substantial uncertainty, and that extrapolations over long periods can easily lead to results that make no sense.

References

- Afriat, Sydney, 1967, "The construction of utility functions from expenditure data," *International Economic Review*, 8(1), 67–77.
- Afriat, Sydney, 1972, "The theory of international comparisons of real income and prices," in D. J. Daly, ed., *International comparisons of prices and output, Proceedings of the conference at York University, Toronto*, New York, National Bureau of Economic Research, 13–84.
- Alterman, W. F., W. Erwin Diewert, and Robert C. Feenstra, 1999, *International trade price indexes and seasonal commodities*, Washington, DC. Bureau of Labor Statistics.
- Asian Development Bank, 2007, *Purchasing power parities and real expenditures*, Manila, December.
- Aten, Bettina, 2006. 'Interarea price levels: an experimental methodology', *Monthly Labor Review*, 129(9), September, 47–61.
- Atkin, David, 2009, "Trade, tastes, and nutrition in India," Department of Economics, Yale University, October. http://www.econ.yale.edu/~da334/Tastes_and_Nutrition.pdf
- Balk, Bert M., 1996, "A comparison of ten methods for multilateral international price and volume comparisons," *Journal of Official Statistics*, 12:2, 199-222.
- Bhalla, Surjit S, 2008, "World Bank—most Asians dead in 1950," *Business Standard*, August 23.
- Brandt, Loren, and Carsten Holz, 2004, "Spatial Price Differences in China: Estimates and Implications," *Economic Development and Cultural Change*, 55(1), 43–86.

- Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers, 2006, "Missing in action: teacher and health worker absence in developing countries," *Journal of Economic Perspectives*, 20(1), 91–116.
- Chen, Shaohua, and Martin Ravallion, 2008, "China is poorer than we thought, but no less successful in the fight against poverty," Washington, D.C., Policy Research Working Paper 4621 (May).
- Ciccone, Antonio and Marek Jarocinski, 2008, "Determinants of economic growth: will data tell?, ICREA-Universitat Pompeu-Fabra, September.
http://www.crei.cat/people/ciccone/growth_determinantswp.pdf
- Deaton, Angus, 2010, "Quality, inequality, and the measurement of global poverty," *American Economic Review*, March. (AEA Presidential Address.)
- Diewert, W. Erwin, 1976, "Exact and superlative index numbers," *Journal of Econometrics*, 4(2), 115–45.
- Diewert, W. Erwin, 1999, "Axiomatic and economic approaches to international comparisons," in Alan Heston and Robert Lipsey, eds., *International and interarea comparisons of income, output, and prices*, University of Chicago Press, 13–52.
- Diewert, W. Erwin, 2001, "The consumer price index and index number purpose," *Journal of Economic and Social Measurement*, 27, 167–248.
- Diewert, W. Erwin, 2007, "Index numbers," in John Eatwell, Murray Milgate, and Peter Newman, eds., *The new Palgrave: a dictionary of economics*, Vol. 2, 767–80.
- Diewert, W. Erwin, 2008a, "New methodology for linking the regions," Dept. of Economics, University of British Columbia (September.) Available at <http://www.econ.ubc.ca/diewert/dp0807.pdf>
- Diewert, W. Erwin, 2008b, "New methodological developments for the International Comparison Program," Dept. of Economics, University of British Columbia (September.) Available at <http://www.econ.ubc.ca/diewert/dp0808.pdf>
- Dikhanov, Yuri, 1997, "Sensitivity of PPP-based income estimates to choice of aggregation procedures," Development Data Group, World Bank, Washington, DC. Available at <http://siteresources.worldbank.org/ICPINT/Resources/icppapertotal.pdf>

- Dowrick, Steve, and John Quiggin, 1994, "International comparisons of living standards and tastes: a revealed preference approach," *American Economic Review*, 84:1, 332–41.
- Economic and Social Commission for Asia and the Pacific (1999), *ESCAP Comparisons of Real Domestic Product and Purchasing Power, 1993*, United Nations, Bangkok.
- Eltető, O., and P. Köves, 1964, "On a problem of index number computation relating to international comparison," *Statisztikai Szemle* 42, 507–18.
- Feenstra, Robert C., Alan Heston, Marcel P. Timmer and Haiyan Deng , 2009, "Estimating Real Production and Expenditures Across Nations: A Proposal for Improving the Penn World Tables", *Review of Economics and Statistics*, 91:1, 201–12.
- Geary, Roy C., 1958, "A note on the comparison of exchange rates and purchasing power between countries," *Journal of the Royal Statistical Society, Series A (General)*, 121(1), 97–99.
- Gershenkron, Alexander, 1947, "The Soviet indices of industrial production," *Review of Economic Studies*, 29(4), 217–26.
- Gini, Corrado, 1924, "Quelques considerations au sujet de la construction des nombres indices des prix et des questions analogues," *Metron*, 4, 3–162.
- Gong, Cathy Honge and Xin Meng (2008) "Regional Price Differences in Urban China 1986-2001: Estimation and Implication," Discussion Paper No. 3621, July, Australian National University and IZA, iza@zia.org.
- Heston, Alan (2008) "The 2005 Global Report on PPPs: A Preliminary Review," *Economic and Political Weekly* (Mumbai), March 15, 2008, 65–9.
- Iklé, Doris M., 1972, "A new approach to the index number problem," *Quarterly Journal of Economics*, 86:2, 188–211.
- Johnson, Simon, William Larson, Chris Papageorgiou and Arvind Subramanian (2009), "Is newer better? Penn World Table revisions and their impact on growth estimates," NBER Working Paper No. 15455, October.
- Khamis, Salem H., 1972, "A new system of index numbers for national and international purposes," *Journal of the Royal Statistical Society, Series A (General)*, 135(1), 96–121.
- Maddison, Angus, 2003, *The world economy: historical statistics*, Paris, OECD.

- Maddison, Angus, 2007, *Chinese Economic Performance in the Long Run: Second Edition, Revised and Updated 960-2030 AD*, Paris, OECD. Appendix D.
- Neary, J. Peter, 2004, "Rationalizing the Penn World Table: true multilateral indexes for international comparisons of real income," *American Economic Review*, 94(5), 1411–28.
- Nuxoll, Daniel A., 1994, "Differences in relative prices and international differences in growth rates," *American Economic Review*, 84(5), 1423–36.
- Organization for Economic Cooperation and Development, 2009, http://stats.oecd.org/Index.aspx?datasetcode=SNA_TABLE4# (downloaded Oct. 25, 2009).
- Pritchett, Lant, 1997, "Divergence, big time," *Journal of Economic Perspectives*, 11(3), 3–17.
- Rogoff, Kenneth, 1996, "The purchasing power parity puzzle," *Journal of Economic Literature*, 34(2), 647–68.
- Seers, Dudley, 1983, *The political economy of nationalism*, Oxford. Oxford University Press.
- Sun, Changqing, and Eric Swanson, 2009, "Estimation of PPPs for non-benchmark economies for the 2005 ICP round," *ICP Bulletin*, 6:1, 20–23.
- Srinivasan, T. N., 1994, "Data base for development analysis: an overview," *Journal of Development Economics*, 44, 3–27.
- Summers, Robert, and Alan Heston, 1991, "The Penn World Table (Mark 5): An expanded set of international comparisons, 1950–1988," *Quarterly Journal of Economics*, 106(2), 327–68.
- Szulc, B., 1964, "Indices for multiregional comparisons," *Przegląd Statystyczny*, 3, 239–54.
- Van Veelen, Matthijs, 2002, "An impossibility theorem concerning multilateral international comparisons of volumes," *Econometrica*, 70, 369–75.
- World Bank, 1993, *Purchasing power of currencies*, prepared for the socioeconomic data division, international economics department.
- World Bank, 2007, *World Development Indicators*, Washington, DC. (CD.)

World Bank, 2008a, *Global purchasing power parities and real expenditures: 2005 international comparison program*, Washington, DC.

World Bank, 2008b, "ICP 2003–2006 Handbook," available online at <http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/ICPEXT/0,contentMDK:20126612~pagePK:60002244~piPK:62002388~theSitePK:270065,00.html>

World Bank, 2008c, *World Development Indicators*, Washington, DC. (CD)

Table 1: Ratios of Laspeyres to Paasche price indexes, USA and Nigeria versus selected other countries

	USA		NIGERIA
<i>Ten smallest</i>			
Ireland	1.048	Congo	0.962
Canada	1.057	Latvia	1.003
Austria	1.062	Guinea	1.030
Switzerland	1.068	Mali	1.032
Germany	1.072	Zimbabwe	1.049
Luxemburg	1.080	Albania	1.050
Italy	1.091	Estonia	1.051
Australia	1.092	Equatorial Guinea	1.053
Belgium	1.094	Burkina Faso	1.055
UK	1.096	Chad	1.062
<i>Ten largest</i>			
Djibouti	2.798	Oman	1.674
Armenia	2.811	Korea	1.707
Moldova	2.905	Japan	1.747
Rwanda	2.943	USA	1.780
Burundi	3.055	Iran	1.781
Zimbabwe	3.771	Kuwait	1.945
Chad	4.033	Bahrain	1.978
Gambia	4.210	Luxemburg	2.095
Kyrgyzstan	5.107	Tajikistan	2.473
Tajikistan	9.615	Qatar	2.847
<i>Other important</i>			
China	1.658	China	1.375
India	1.611	India	1.390
Indonesia	1.484	Indonesia	1.405
Brazil	1.508	Brazil	1.329
Nigeria	1.780	South Africa	1.069
Russia	1.823	Russia	1.092

Table 2: Bilateral Fisher, GEKS, Geary-Khamis, and Iklé-Dikhanov-Balk indexes, selected countries as in Table 1

	LP-Spread	Bilateral Fisher	GEKS	Geary-Khamis	IDB	ICP 2005
<i>Ten smallest</i>						
Ireland	1.048	1.042	1.023	1.100	1.143	1.023
Canada	1.057	1.195	1.176	1.237	1.265	1.214
Austria	1.062	0.878	0.852	0.902	0.928	0.874
Switzerland	1.068	1.748	1.709	1.695	1.747	1.741
Germany	1.072	0.896	0.860	0.873	0.904	0.893
Luxemburg	1.080	0.938	0.924	0.854	0.884	0.922
Italy	1.091	0.884	0.853	0.890	0.908	0.875
Australia	1.092	1.375	1.319	1.401	1.413	1.388
Belgium	1.094	0.889	0.880	0.889	0.942	0.899
UK	1.096	0.637	0.649	0.639	0.648	0.649
<i>Ten largest</i>						
Djibouti	2.798	73.62	84.69	65.94	86.30	84.69
Armenia	2.811	145.15	147.07	123.53	151.35	178.58
Moldova	2.905	3.617	3.513	3.050	3.689	4.434
Rwanda	2.943	176.28	188.63	152.94	195.87	186.18
Burundi	3.055	315.61	343.34	276.49	360.99	342.96
Zimbabwe	3.771	26702	30671	21900	29535	33068
Chad	4.033	163.04	205.92	143.0	206.95	208.0
Gambia	4.210	6.315	8.019	5.530	8.041	7.560
Kyrgyzstan	5.107	8.157	9.068	6.390	8.404	11.354
Tajikistan	9.615	0.457	0.542	0.296	0.425	0.744
<i>Other important</i>						
China	1.658	3.127	3.059	2.941	3.290	3.448
India	1.611	14.755	13.480	13.605	14.757	14.669
Indonesia	1.484	3833	3606	3540	3867	3934
Brazil	1.508	1.346	1.376	1.401	1.577	1.357
Nigeria	1.780	54.77	57.42	54.38	65.31	60.23
Russia	1.823	11.314	11.163	11.397	12.978	12.736

Notes: US is the base country. LP-spread is the Laspeyres-Paasche spread as in Table 1. Bilateral Fisher is the Fisher price index for each country relative to the US, calculated as a bilateral comparison. EKS, Geary-Khamis, and Iklé-Dikhanov-Balk are the multilateral PPP indexes for GDP excluding the trade balance. ICP2005 is the PPP from the WDI 2008.

Source: Authors' calculations using basic heading parities from the 2005 ICP.

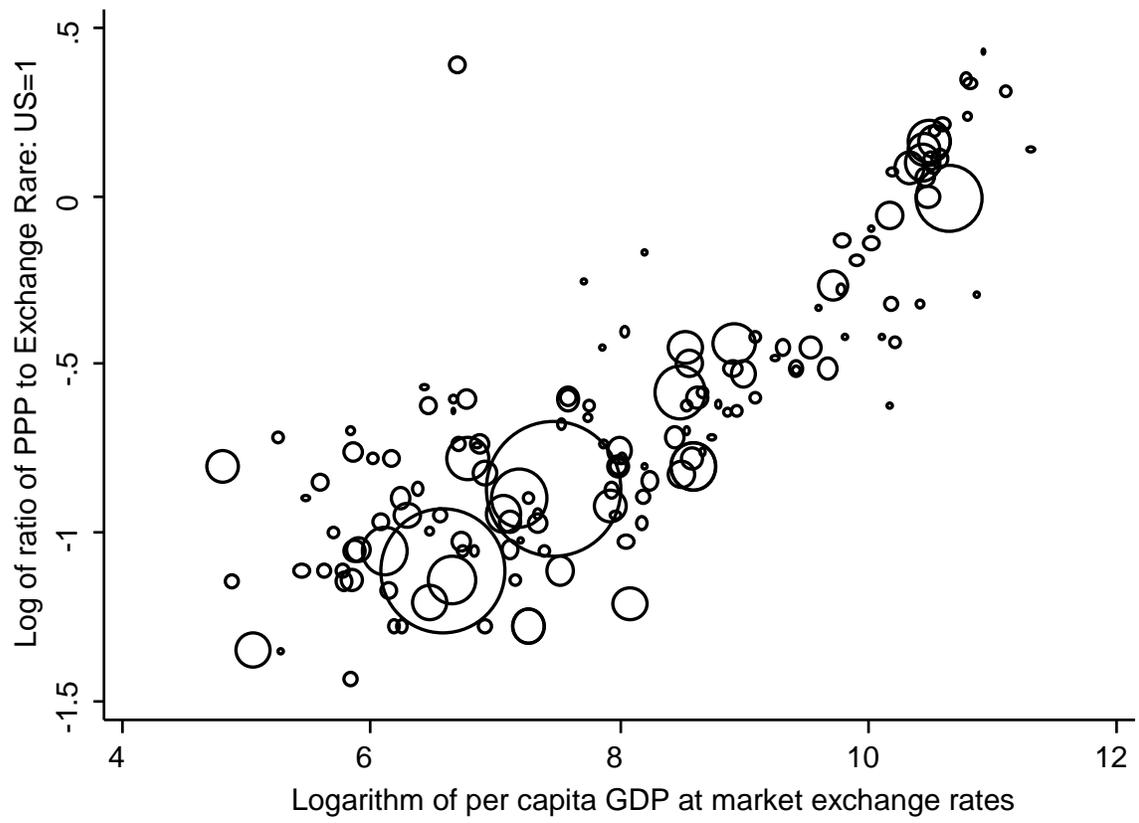


Figure 1: Price levels in relation to GDP

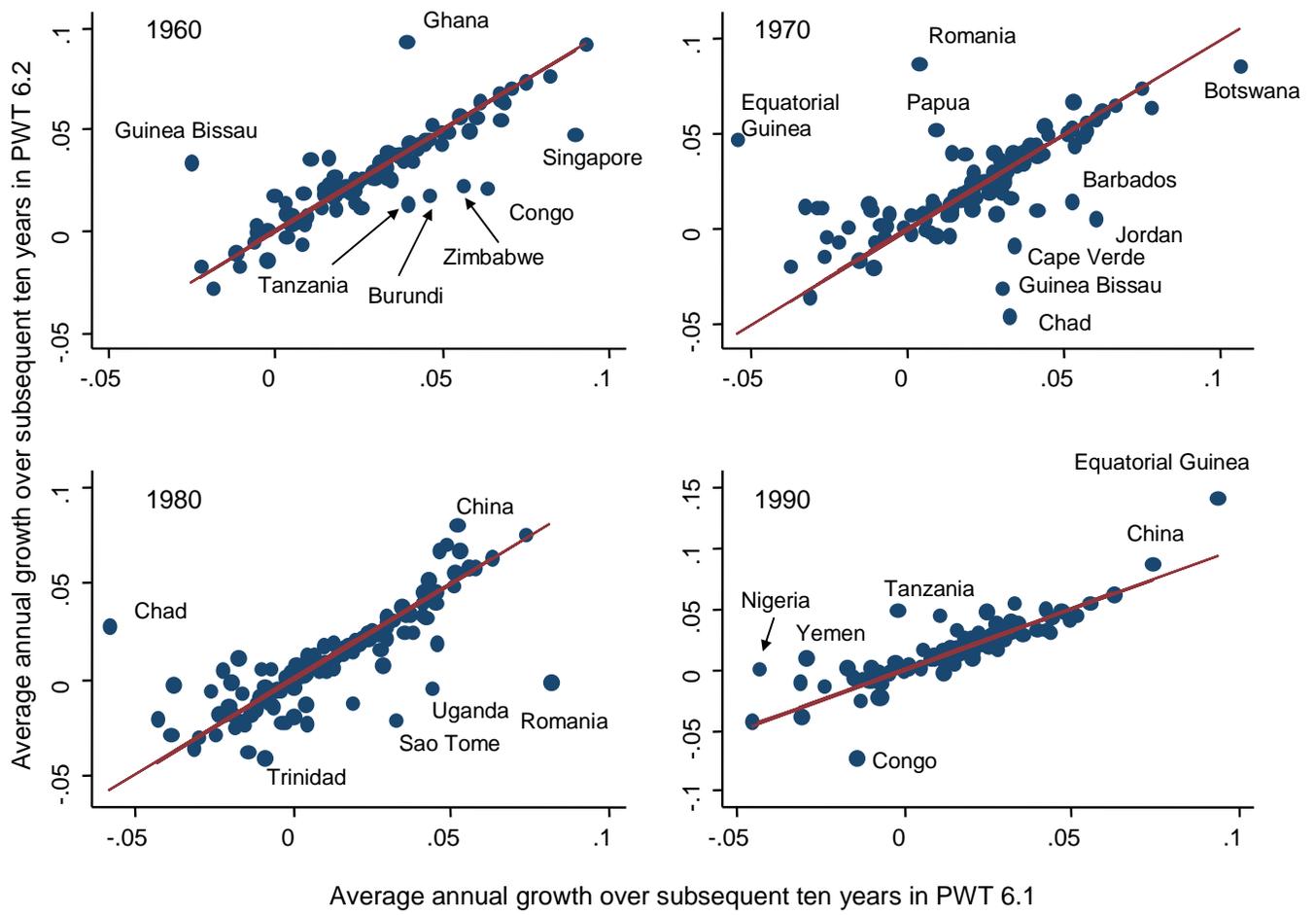


Figure 2: Revisions to ten year growth rates in PWT6.1 and PWT6.2