Comment on Understanding PPPs and PPP Based National Accounts

Erwin Diewert,[1]
Discussion Paper 10-05, March 16, 2010
Department of Economics,
University of British Columbia,
Vancouver, Canada, V6T 1Z1.
Email: diewert@econ.ubc.ca

Abstract

The paper discusses the three major problems that Deaton and Heston identify that make it difficult to construct consistent estimates of real output across countries: (i) the choice of a multilateral index number formula; (ii) weak national accounts estimates of country expenditures on the 155 major expenditure categories that are identified in the World Bank's International Comparison Program and (iii) the fact that countries at different levels of development consume products that are not consumed widely by other countries in the comparison and thus the country price levels that are determined in the comparison project may not be reliable.

Journal of Economic Literature Classification Numbers

C43, C81, E31, O57.

Keywords

Index numbers, multilateral comparison methods, ICP 2005, GEKS, EKS, Geary, Khamis, Balk, Dikhanov, Iklé, Country Product Dummy (CPD) method, basic headings, Purchasing Price Parities (PPPs), spatial chaining, fixity.

[1] The author is indebted to Steven Davis, Angus Deaton and Alan Heston for helpful comments and to the SSHRC of Canada for financial support. This paper is an expanded version of a shorter comment which will appear in the *American Economic Journal*: *Macroeconomics*.

1. Introduction

Angus Deaton and Alan Heston (2010) have written an excellent paper that explains how the latest round of the World Bank's 2005 International Comparison Program (ICP) was conducted and how it will affect the Penn World Tables in the near future. In brief, the 2005 round of the ICP compared relative price levels and GDP levels across 146 countries. In this round of the ICP, the world was divided into 6 regions: OECD, CIS, Africa, South America, Asia Pacific and West Asia. What was new in this round compared to previous rounds of the ICP was that each region was allowed to develop its own product list and collect prices on this list for countries in the region. The regions were then linked using another separate product list and 18 countries across the 6 regions collected prices for products on this list and this information was used to link prices and quantities across the regions. An additional complication was that the final linking of prices and volumes (or quantities) across regions had to respect the regional price and volume measures that were (separately) constructed by the regions.[2] As Deaton and Heston indicate, there is no doubt that the quality of the resulting international comparisons was greatly improved by these new methodological developments as compared to previous rounds. But as Deaton and Heston also indicate, there are still some significant problems with the 2005 comparisons, which they address in their paper and which we will review to some extent in this comment.

Before we address some of the general problems associated with making international comparisons of real output and price levels, it is useful to discuss why such comparisons are of interest to macroeconomists. Economists are interested in why some countries grow faster than other countries and they come up with theories that attempt to explain these differential growth rates. Economists are also interested in whether rich countries grow more slowly than less developed countries; i.e., they are interested in whether levels of per capita output are converging or not. However, in order to test alternative theories of growth, it is first necessary to *measure* levels of real output across countries in comparable units. As Deaton and Heston explain in their paper, if the law of one price were true (i.e., international trade equalizes exchange rate adjusted prices across all countries for the same item), then we could simply use market exchange rates to deflate

[2] This is the *regional fixity problem*. For more details on the new ICP methodology, see the World Bank (2007) (2008) and Diewert (2010a).

the nominal GDP of each country into a meaningful real GDP which would be comparable across countries. But of course, the law of one price does not hold for a number of reasons: (i) some firms have some market power and they price discriminate across countries; (ii) many services are not internationally traded; (iii) transport costs lead to different prices across countries for the same homogeneous product and finally, (iv) capital movements are huge and often lead to large stochastic movements in exchange rates. Thus there is a need to directly collect prices for identical products across countries in order to compare real output across countries.

Deaton and Heston (2010) identify three major problems that make it difficult to construct consistent estimates of real output across countries:

- How exactly should we choose an *index number formula* that will enable us to compare real outputs across countries? Because country expenditure shares and relative prices by commodity category can differ much more in the cross sectional context than in the time series context, different index number formulae can give very different estimates of the relative size of country outputs whereas the problem is not as severe in the time series context where there is more stability in expenditure shares and relative prices across time periods. There is also a related problem that while index number theory is quite well developed in the context of making bilateral comparisons (comparisons between two countries or two time periods), there is less consensus on what the appropriate formula should be in the multilateral case.[3]
- Index number comparisons between countries require information on the prices of products in the same units of measurement and information on *expenditure shares* by product category. The ICP gets its country information on expenditure shares (by 155 major product categories) by using information on final demand expenditures from the *national accounts* of the participating countries. However, these national expenditure estimates are not always very accurate or constructed on a consistent basis across countries and these inaccurate estimates lead to poor estimates of real output across countries in some cases.
- The final problem that Deaton and Heston flag is that it is very difficult or impossible to compare real outputs across countries that are at very different stages of development, since poor countries may be consuming products which are very different from the products being consumed in rich countries; i.e., it is *impossible to compare the incomparable*.

In the following two sections of this comment, I will discuss the problem of choosing an index number formula in more detail while sections 4 and 5 will discuss the last two difficulties listed above.

2. Alternative Multilateral Index Number Formulae

[3] There is a growing consensus that a superlative index such as the Fisher ideal index or the Törnqvist index is an appropriate bilateral index from many points of view; see Diewert (1976) (1978) and the ILO (2004). However, there is less consensus on what the appropriate multilateral formula should be; see Balk (1996) and Diewert (1999) for a review of the properties of many multilateral formulae.

The index number difficulties associated with making comparisons of real output across countries can be grouped into two separate categories:

- Comparisons of prices across countries below the Basic Heading level, where expenditure information on the commodity categories is not available and
- Comparisons of prices and expenditures at the Basic Heading level where we have country expenditure and price information.

In the ICP, there were 155 Basic Heading expenditure categories. Within each of these expenditure categories, there was no quantity or expenditure data available across countries to associate with the prices of the individual products that were priced across countries. Deaton and Heston do a good job of explaining the ICP methodology at this first level of aggregation where expenditure or quantity information is not available. Additional details on this first stage methodology can be found in the World Bank (2007) and in Diewert (2004) (2010a). I will only note that the methodology for linking the regions will be different for the next round of the International Comparison Program in 2011. In ICP 2005, the regions were linked by the modification of Summer's (1973) *Country Product Dummy* method that was described in Deaton and Heston. This method required the collection of prices across 18 *ring countries* in the various regions, using a separate *core list of products* that was different from the *regional product lists*. In ICP 2011, a new list of core products will be constructed and this core list will be added to the separate regional list of products to be priced within each region, and every country in the comparison will be instructed to price as many of the core list products that are available (and representative) in their respective countries. Thus it is hoped that the linking of regional prices at the Basic Heading level in ICP 2011 will be more reliable than was the case for ICP 2005, where the interregional linking process perhaps depended too much on just a few countries.

We turn now to the problems associated with the linking of countries at the Basic Heading level of aggregation, where price and expenditure information for 155 product categories was available for each country in the comparison. As Deaton and Heston (2010) mention, there are *three main multilateral index number methods* in use in the ICP and the Penn World Tables:

- GEKS, the multilateral method based on Fisher (1922) ideal index bilateral comparisons, due to Gini (1924), Eltetö and Köves (1964) and Szulc (1964);
- GK, the Geary (1958) and Khamis (1972) method and
- IDB, the Iklé, Dikhanov and Balk method.[4]

[4] Iklé (1972; 203) proposed the equations for the method in a rather difficult to interpret manner and provided a proof for the existence of a solution for the case of two countries. Dikhanov (1994; 6-9) used a much more transparent set of equations, explained the advantages of the method over the GK method and illustrated the method with an extensive set of computations. Balk (1996; 207-208) used the Dikhanov equations and provided a proof of the existence of a solution to the system for an arbitrary number of countries. Diewert (2010a) provided an alternative set of equations to characterize the method and looked at the properties of the method. The properties of these and other multilateral methods are discussed in Balk (1996), Hill (1997) and Diewert (1999) (2010a).

It should be noted that the GK and IDB methods are both *additive methods*; i.e., using these methods, the real output of any country is proportional to a price weighted sum of the individual quantity or volume components of the country's GDP, where the price weights are common across all countries. The use of an additive method is tremendously useful for many analytical purposes, because with an appropriate choice of units, we can construct a table of real country expenditures on the 155 commodity classes where the components will nicely add up across the entries in a row to give aggregate real output for that country and the entries will add up down a column to give aggregate world final demand for that product category. Unfortunately, as I will indicate later, additivity is not consistent with normal substitution behavior on the part of final demanders (while GEKS is consistent) and thus for some purposes (such as the measurement of welfare and poverty), an additive approach is not appropriate. Thus I agree with Deaton and Heston that the ICP needs to provide users with at least two sets of results, one of which is an additive method (and it appears that IDB is "better" than GK for the reasons indicated by Deaton and Heston) and one of which is based on an economic approach to international comparisons such as GEKS.

In ICP 2005, aggregate PPPs and relative country volumes for countries within each region were constructed for five of the six regions using the GEKS method. However, the African region wanted to use an additive method and so this region used the IDB method for constructing PPPs and relative volumes within the region.

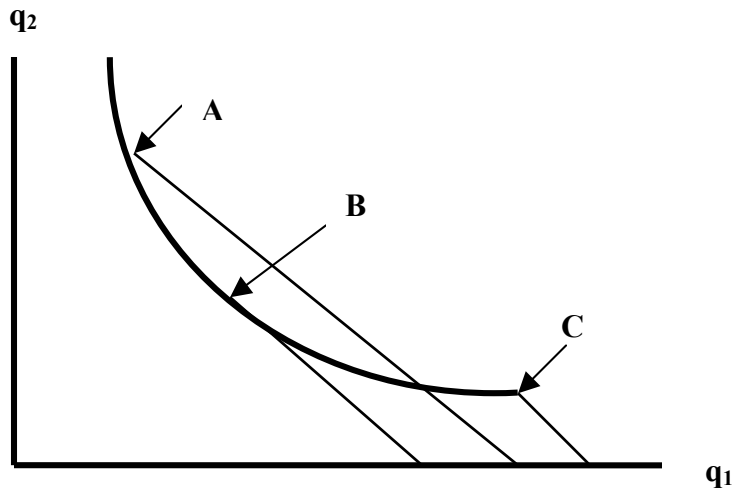
Once the PPPs and real GDP estimates for each country in a region have been constructed (using either the GEKS or IDB method as indicated above for the regions), we then encounter the *fixity problem*; i.e., for political reasons, the estimates within each region have to be linked across regions without affecting the PPPs and relative country volumes within each region. Deaton and Heston describe how this was done using a method suggested by Diewert (2004). Sergeev (2009b) noted a drawback of this method: although the method does not depend on the choice of the numeraire region, it does depend on the choice of the numeraire country within each region. This aspect of the interregional linking method is not satisfactory and so for ICP 2011, a better solution to the interregional fixity problem will have to be found. At this stage, it is an open question on what method will be used to link the regions.

There is a tension between additive multilateral methods (such as GK and IDB) and methods that are consistent with superlative indexes (such as GEKS). Recall that an additive method is a method that calculates the real GDP of a country by adding up its component volumes or quantities using a vector of (price) weights that is constant across countries in the comparison. However, an additive method is not consistent with an *economic approach* to index number theory[5] if the number of countries being compared is greater than two. This last point can be explained with the help of a diagram.[6]

[5] An economic approach to index number theory is one where purchasers of the goods and services in the domain of definition of the index have preferences over different combinations of the commodities; e.g., see Diewert (1976).

[6] This diagram is basically due to Marris (1984; 52) and Diewert (1999; 48-50).

Figure 1



The solid curved line in the above Figure represents an indifference curve for purchasers of the two goods under consideration. The consumption vectors of Countries A, B and C are all on the same indifference curve and hence, the multilateral method should show the same volume for the three countries. If we use the relative prices that country B faces as "world" reference prices in an additive method, then country B has the lowest volume or real consumption, followed by country A and finally, C has the highest volume. But they all have equal volumes! It can be seen that we can devise an additive method that will make the volumes of any two countries equal but we cannot devise an additive method that will equalize the volumes for all three countries. On the other hand, the common indifference curve in Figure 1 can be approximated reasonably well by a flexible functional form that has a corresponding exact index number formula (such as the Fisher or Törnqvist indexes) and thus a GEKS method that used either of these bilateral indexes as a basic building block would give the right answer to a reasonable degree of approximation.[7]

[7] Deaton and Heston (2010) explain how to derive the GEKS multilateral system using a least squares minimization problem, due originally to Gini (1924), that will essentially make an M by M matrix of bilateral Fisher parities that are not transitive into a best fitting set of transitive parities. Another well known method for deriving the GEKS parities works as follows. Pick any country as the base country and use the Fisher bilateral quantity index to form the real output of every country relative to the chosen base country. This gives us estimated volumes for all countries in the comparison relative to the chosen base country. Now repeat this process, choosing each country in turn as the base country, which gives us M sets of relative volume estimates. The final step for obtaining the GEKS relative volumes is to take the geometric mean of all of the M base country dependent sets of parities. If final demanders in each country had identical homothetic preferences that satisfied the functional form assumptions listed in Diewert (1976), then the base country dependent parities would all be proportional to each other, a condition which does not hold in practice.

Deaton and Heston mention in passing Neary's (2004) GAIA multilateral system, which they describe as a consumer theory consistent version of the GK system, which allows for nonhomothetic preferences on the part of final demanders. Deaton and Heston also point out that a weakness of the Neary multilateral system 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 in each country. This problem was also noticed by Feenstra, Ma and Rao (2009) and these authors generalized Neary's framework to work with two sets of cross sectional data in order to estimate preferences[8] and they also experimented with alternative sets of reference prices. Barnett, Diewert and Zellner (2009) in their discussion of Feenstra, Mao and Rao, noted that a natural generalization of their model would be to use a set of reference prices which would be representative for each country in the comparison. Using representative prices for each country would lead to M sets of relative volumes and in the end, these country specific parities could be averaged just as the GEKS method averages country specific parities. Barnett, Diewert and Zellner conjectured that this geometric average of the country estimates will probably be close to GEKS or Caves, Christensen and Diewert (1982) estimates based on traditional multilateral index number theory, which of course, does not use econometrics.

3. Spatial Comparisons Based on Similar Price Structures

The GEKS multilateral method treats each country's parity as being equally valid and hence an averaging of the parities is appropriate under this hypothesis. Thus the method is "democratic" in that each bilateral index number comparison between any two countries gets the same weight in the overall method. However, it is not the case that all bilateral comparisons of volume between two countries are equally accurate: if the relative prices in countries A and B are very similar, then the Laspeyres and Paasche quantity indexes will be very close to each other and hence it is likely that the "true" volume comparison between these two countries (using the economic approach to index number theory) will be very close to the Fisher volume comparison. On the other hand, if the structure of relative prices in the two countries is very different, then it is likely that the structure of relative quantities in the two countries will also be different and hence the Laspeyres and Paasche quantity indexes will likely differ considerably and we can no longer be certain that the Fisher quantity index will be close to the "true" volume comparison. The above considerations suggest that a more accurate set of world product shares could be constructed if we started out making a bilateral comparison between the two countries which had the most *similar relative price structures*. We could then look for a third country which had the most similar relative price structure to the first two countries and link in this third country to the comparisons of volume between the first two countries and so on. At the end of this procedure, we would have a *minimum spanning tree*: a path between all countries that minimized the sum of the relative price similarity measures. This linking methodology has been developed by Robert Hill (1999a) (1999b) (2004) (2009).[9] The bottom line is that spatial linking using Fisher ideal

[8] Methods that rely on the econometric estimation of preferences across countries are probably not suitable for the ICP, since it becomes very difficult to estimate flexible preferences for 155 commodity categories.

[9] A key aspect of this methodology is the choice of the measure of similarity (or dissimilarity) of the relative price structures of two countries. Various measures of the similarity or dissimilarity of relative

quantity indexes as the bilateral links is an alternative to GEKS which has some advantages over it.[10] Both methods are consistent with the economic approach to index number theory.

The narrowing of Paasche and Laspeyres spreads by the use of a spatial linking method is not the only advantage of this method of linking countries. There are advantages at *lower levels of aggregation* in that if we compare similar in structure countries, we will find that product overlaps are maximized:

"Many differences in quality and proportion of high tech items discussed above are likely to be more pronounced between countries with very different economic structures. If criteria can be developed to identify countries with similar economic structure and they are compared only with each other, then it may overcome many of the issues of quality and lowest common denominator item comparisons. Economically similar countries are likely to have outlet types in similar proportions carrying the same types of goods and services. So direct comparisons between such countries will do a better job of holding constant the quality of the items than comparisons across more diverse countries." Bettina Aten and Alan Heston (2009; 251).

The above quotation suggests that perhaps the similarity criterion should not be based only on the similarity of the structure of relative prices across the two countries being compared. In addition, we should look at the degree of similarity in the structure of absolute per capita quantity vectors and take a sum of the two measures of similarity as our overall measure of similarity in structure.

There are some disadvantages to the spatial linking method. The two most important disadvantages are:

- The path of bilateral links between countries generated by the method tends to be unstable; i.e., the most similar tree linking the countries tends to change when we move from one cross sectional comparison between countries to another cross sectional comparison.
- Some countries in the comparison will inevitably have lower quality data than other countries and if these poorer data quality countries end up having many bilateral links with many countries in the minimum spanning tree, then the quality of the entire comparison may be low.

Hill (2009) discusses both of these problems and offers "reasonable" solutions to these difficulties. The first difficulty is not really a difficulty if the overall volume comparisons

---

price structures have been proposed by Aten and Heston (2009), Diewert (2009a), Hill (1997) (2009) and Sergeev (2009a).

[10] Deaton (2010; 33-34) noticed the following problem with the GEKS method: suppose we have two countries where the expenditure share on commodity 1 is tiny for country A and very big for country B. Suppose also that the price of commodity 1 in country A is very large relative to the price in country B. Then looking at the Törnqvist price index between A and B, it can be seen that the overall price level for country A will be blown up by the relatively high price for good 1 in A relative to B and by the big expenditure share in B on commodity 1. Since the Törnqvist will generally approximate the corresponding Fisher index closely, it can be seen that we have ended up exaggerating the price level of country A relative to B. This problem can be mitigated by spatial linking of countries that have similar price and quantity structures.

remain more or less the same even if the particular bilateral links change. In particular, it may be the case that countries break up into two or more relatively homogeneous groups. Within each group, the bilateral dissimilarity measures are all low so even if the links within each group change, the relative volume indexes within each group remain roughly unchanged. The key problem then boils down to the bilateral links between the various groupings. In order to get more stability between these groupings, it may be advisable to have more than one link between the groupings and this constraint can readily be imposed. The second difficulty can be dealt with by specifying that countries with lower quality data should not be allowed to have more than one link in the overall tree of comparisons.

Of course, a problem with the above "solutions" to the problems associated with spatial linking is that the solutions appear to have an ad hoc character and this may lead to charges by outside observers that the ICP is being manipulated. This potential problem could be mitigated by experimentation with the ICP 2005 data set so a firm a priori strategy could be put in place before the results for ICP 2011 were calculated.

4. Problems Associated with the National Accounts Expenditure Estimates

Deaton and Heston (2010) rightly stress that the overall accuracy of the real product estimates across countries and time that are based on the ICP 2005 estimates are adversely affected by inaccurate estimates of expenditures on the 155 Basic Heading national accounts categories by country. Some of the inaccuracies are due to the fact that the participating countries do not always have a resource base that is adequate to form reasonably accurate estimates of expenditures. The World Bank and the various agencies involved in ICP 2011 are aware of these deficiencies and are attempting to provide resources and expert help to improve the quality of country national accounts for the 2011 round. However, I would like to flag *three areas* where all countries, rich and poor, are generally having difficulties in measuring expenditures:

- Expenditures on housing services, particularly the services of Owner Occupied Housing (OOH);
- Expenditures on financial services such as banking and insurance services and
- Expenditures on government goods and services.

This is not the appropriate place to explain all of the conceptual difficulties that are associated with economic measurement in the above problem areas. However, I would like to give interested readers a few recent references that get into some of the problems associated with each area. For problems associated with the measurement of housing services, see Diewert (2009b) (2009c), Diewert and Nakamura (2009), Diewert, Nakamura and Nakamura (2009) and Garner and Verbrugge (2009a) (2009b). For problems associated with the measurement of financial services, see Fixler (2009) and Wang, Basu and Fernald (2009). Finally for problems associated with measuring government outputs, see Atkinson (2005), Diewert (2009) (2010b) and Schreyer (2009). In general, the conceptual problems associated with economic measurement in each of the above areas is are quite complex. It would be desirable for more economists to get

interested in these measurement problems and it would be useful for more economists to interact with statistical agency researchers to ensure that these agencies are producing numbers that are at least somewhat consistent with current economic measurement concepts.

Deaton and Heston (2010) note that successive ICP rounds, when combined with country time series data on real GDP, are generally not consistent with each other. The following list of reasons may help to explain these discrepancies:

- The time series estimates of real GDP by country are inherently more accurate than the corresponding cross sectional estimates because the structure of relative prices and quantities over time is very similar in the time series context (and so all reasonable indexes will tend to agree) whereas the structure of relative prices and quantities is very different across countries and so the cross country comparisons will be inherently less reliable. The problems of making valid cross country comparisons are intensified by ICP methodology at the product level, which in the end, will price products which are not representative for many countries (but are still available). This will lead to cross country Basic Heading parities which have substantial measurement errors in them.
- As indicated above, some countries will be able to construct national accounts expenditure estimates by category which are more reliable than other countries. The degree of reliability will tend to change over time, leading to variable cross country comparisons. For the difficult to measure sectors of the economy mentioned above, the estimated expenditures for these categories may have a huge stochastic element due to inadequate methodology.
- In the ICP, the country PPPs for exports and imports are simply country average market exchange rates for the reference year. But when a country constructs its own estimates of real GDP, there is always an attempt to deflate the value of exports by and export price index and the value of imports by an import price index. Thus changes in the terms of trade for a country will affect its time series estimates of real GDP but changes in the terms of trade will not affect the cross sectional comparisons of real GDP. Since international trade as a share of the economy has generally grown rapidly over the past 60 years and since there are dramatic changes in the terms of trade across countries, this ICP treatment of international trade could explain some of the discrepancy between the time series and cross sectional estimates of real GDP.

5. Comparing the Incomparable

Deaton and Heston (2010) raise the issue of the impossibility of making comparisons across countries when the commodity bundles being consumed in the two countries are entirely different.[11] Deaton in a communication with the present author suggested the following numerical example where we have only two commodities and three countries:

[11] Hill and Timmer (2006) deal with a less extreme situation where there is some matching across commodity groups and they attempt to solve it via a statistical solution. But their suggested solution is not altogether convincing.

Table 1: Prices for 3 Countries and 2 Commodities

Good	Country 1	Country 2	Country 3
1	p_{11}	p_{21}	•
2	p_{12}	•	p_{32}

The price p_{cn} is the price of good n in country c, and a • indicates that the corresponding commodity is not consumed at all in the country and thus there is no price. Deaton further supposed for simplicity that in country 1, half the budget gets devoted to each of the two goods, so that we do not have to worry too much about weights for this example. Deaton then goes on to explain that there is really no basis for comparing country 2 and 3 since there is no commodity overlap between them. Hence he is rightly suspicious of any method which "solves" this problem, since any solution is arbitrary.

There are at least two methods we could apply to the above data: spatial linking or the use of the Country Product Dummy method due to Summers (1973). Using spatial linking, country 2 is most similar to country 1 and so country 2 would be linked to country 1 via the relative price of commodity 1. Country 3 is also most similar in price structure to country 1 so country 3 would be linked to country 1 via the relative price of commodity 2. Thus the price level of country 1 will be set equal to:

(1) $P^{11} = 1 \equiv \alpha_1$.

The price level of country 2 relative to country 1 will be set equal to:

(2) $P^{21} \equiv p_{21}/p_{11} \equiv \alpha_2$.

And finally, the price level of country 3 relative to country 1 will be set equal to:

(3) $P^{31} \equiv p_{32}/p_{12} \equiv \alpha_3$.

Thus the price level of country 3 relative to country 2 will be

(4) $P^{32} \equiv P^{31}/P^{21} = [p_{32}/p_{12}]/[p_{21}/p_{11}]$.

Thus the minimum spanning tree methodology seems to give us a definite estimate of the price level in country 2 relative to 3 even though at first glance, it appears to be impossible to compare the outputs of countries 2 and 3.

However, there are possible responses to the above Deaton critique. I will present three ways that could be used to justify the rather indirect methodology that emerges from the spatial linking methodology.

Attempt 1: Suppose final demanders in all 3 countries *had the same linear indifference curve preferences*; i.e., suppose that $u = \beta_1 q_1 + \beta_2 q_2$ where β_1 equals p_{11} and β_2 equals p_{12}.

Then the above solution would in fact give us exactly the right answer in terms of welfare! It could be argued that this is an extreme assumption but perhaps it gives us an adequate approximation to the "truth".

Attempt 2: We can justify the above solution as being the same solution that the *country product dummy method* generates, which is a purely "statistical" model. The model works as follows: we set the country c price for a product n, p_{cn} equal to the product of a *country general price level* α_c times a *product effect* β_n. Thus in the present situation, we can fit the data exactly and the estimating equations are as follows:

(5) $p_{11} = \alpha_1\beta_1$;
(6) $p_{12} = \alpha_1\beta_2$;
(7) $p_{21} = \alpha_2\beta_1$;
(8) $p_{32} = \alpha_3\beta_2$.

We also require a normalization on the country general price levels so we choose country 1 as the numeraire country and set

(9) $\alpha_1 = 1$.

We have 5 equations in 5 unknowns and the CPD solution is as follows:

(10) $\alpha_2 = p_{21}/p_{11}$; $\alpha_3 = p_{32}/p_{12}$; $\beta_1 = p_{11}$; $\beta_2 = p_{12}$

which is the same solution as we obtained using spatial linking. Now the CPD method was originally justified as a purely statistical method. But the CPD parities can be given a deeper meaning in terms of *imputed prices for the missing items* in countries 2 and 3. In this particularly simple case, it can be seen that if a final demander in country 2 wanted to purchase a unit of commodity 2, then he or she could purchase a unit of commodity 2 in country 1 at the country 2 price (in terms of units of commodity 1 foregone) of $\alpha_2\beta_2 = [p_{21}/p_{11}]p_{12}$. Similarly, if a final demander in country 3 wanted to purchase a unit of commodity 1, then he or she could purchase a unit of commodity 1 in country 1 at the country 3 price (in terms of units of commodity 2 foregone) of $\alpha_3\beta_1 = [p_{32}/p_{12}]p_{11}$. In fact, this is why the CPD method was invented in the first place by Summers (1973): as a way of filling in missing prices.

Attempt 3: The World Bank and many academic and government economists want to have estimates of relative output quantities and relative price levels across countries. Thus it seems to me that we should do our best to meet this demand! In other words, even though the methods are not perfect, we need to give the World Bank our best advice in meeting this demand. Thus, we have to pick the "best" method for making international comparisons out of the galaxy of possible methods. There is substantial overlap across most countries for most commodities. Thus linking countries in a chain of bilateral comparisons where relative prices and per capita quantities match up as much as possible seems to me to be the "best" method that we have available to us since it tries to make

comparisons when they are possible and match items and categories according to the similarity of price and (per capita) quantity structures across countries.

But in the end, the cautious tone that Deaton and Heston take towards international comparisons is warranted. There are many sources of error and when we compare the real GDP of a pair of countries that are at very different levels of development, the comparison is going to be subject to a substantial amount of uncertainty.

References

Atkinson, Tony (2005), *Atkinson Review: Final Report; Measurement of Government Output and Productivity for the National Accounts*, New York: Palgrave Macmillan.

Aten, B. and A. Heston (2009), "Chaining Methods for International Real Product and Purchasing Power Comparisons: Issues and Alternatives", pp. 245-273 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S. Prasada Rao (ed.), Cheltenham UK: Edward Elgar.

Balk, B.M. (1996), "A Comparison of Ten Methods for Multilateral International Price and Volume Comparisons", *Journal of Official Statistics* 12, 199-222.

Barnett, W.A., W.E. Diewert and A. Zellner (2009), "Introduction to Measurement with Theory", *Macroeconomic Dynamics* 13, Supplement 2, 151-168.

Caves, D.W., L.R. Christensen and W.E. Diewert (1982), Multilateral comparisons of output, input and productivity using superlative index numbers. *Economic Journal* 92, 73-86.

Deaton, A. (2010), "Price Indexes, Inequality and the Measurement of World Poverty", Presidential Address, American Economic Association, Atlanta, January 2010.

Deaton, A. and A. Heston (2010), "Understanding PPPs and PPP Based National Accounts", forthcoming in the *American Economics Journal: Macroeconomics*.

Diewert, W.E. (1976), "Exact and Superlative Index Numbers", *Journal of Econometrics* 4, 114-145.

Diewert, W.E. (1978), "Superlative Index Numbers and Consistency in Aggregation", *Econometrica* 46, 883-900.

Diewert, W.E. (1999), "Axiomatic and Economic Approaches to International Comparisons", pp. 13-87 in *International and Interarea Comparisons of Income, Output and Prices*, A. Heston and R.E. Lipsey (eds.), Studies in Income and Wealth, Volume 61, Chicago: The University of Chicago Press.

Diewert, W.E. (2004), "On the Stochastic Approach to Linking the Regions in the ICP", Discussion Paper 04-16, Department of Economics, The University of British Columbia, Vancouver Canada, V6T 1Z1.

Diewert, W.E. (2008), "The Measurement of Nonmarket Sector Outputs and Inputs using Cost Weights", Discussion Paper 08-03, Department of Economics, University of British Columbia, Vancouver, Canada.

Diewert, W.E. (2009a), "Similarity Indexes and Criteria for Spatial Linking", pp. 183-216 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S. Prasada Rao (ed.), Cheltenham UK: Edward Elgar.

Diewert, W.E. (2009b), "The Paris OECD-IMF Workshop on Real Estate Price Indexes: Conclusions and Future Directions", pp. 87-116 in *Price and Productivity Measurement, Volume 1: Housing*, W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds.), Trafford Press.

Diewert, W.E. (2009c), "Durables and Owner Occupied Housing in a Consumer Price Index", pp. 445-500 in *Price Index Concepts and Measurement*, Studies in Income and Wealth 70, Chicago: The University of Chicago Press.

Diewert, W.E. (2010a), "New Methodological Developments for the International Comparison Program", *Review of Income and Wealth*, forthcoming.

Diewert, W.E. (2010b), "Measuring Productivity in the Public Sector: Some Conceptual Problems", Discussion Paper 10-04, Department of Economics, University of British Columbia, Vancouver, Canada, V6T 1Z1.

Diewert, W.E. and A.O. Nakamura (2009), "Accounting for Housing in a CPI", pp.7-32 in *Price and Productivity Measurement, Volume 1: Housing*, W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds.), Trafford Press.

Diewert, W.E., A.O. Nakamura and L.I. Nakamura (2009), "The Housing Bubble and a New Approach to Accounting for Housing in A CPI", *Journal of Housing Economics* 18, 156-171.

Dikhanov, Y. (1994), "Sensitivity of PPP-Based Income Estimates to Choice of Aggregation Procedures", The World Bank, Washington D.C., June 10, paper presented at 23rd General Conference of the International Association for Research in Income and Wealth, St. Andrews, Canada, August 21-27, 1994.

Eltetö, O. and Köves, P. (1964), "On a Problem of Index Number Computation relating to international comparison", *Statisztikai Szemle* 42, 507-18.

Feenstra, R.C., H. Ma and D.S. Prasada Rao (2009), "Consistent Comparisons of Real Incomes Across Time and Space", *Macroeconomic Dynamics* 13, Supplement 2, 169-193.

Fisher, I. (1922), *The Making of Index Numbers*, Boston: Houghton Mifflin.

Fixler, D. (2009), "Incorporating Financial Services in a Consumer Price Index", pp. 239-271 in *Price Index Concepts and Measurement*, Studies in Income and Wealth 70, Chicago: The University of Chicago Press.

Garner, T.I. and R. Verbrugge (2009a), "The Puzzling Divergence of Rents and User Costs, 1980-2004: Summary and Extensions", pp. 125-146 in *Price and Productivity Measurement, Volume 1: Housing*, W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds.), Trafford Press.

Garner, T.I. and R. Verbrugge (2009b), "Reconciling User Costs and Rental Equivalence: Evidence from the US Consumer Expenditure Survey", *Journal of Housing Economics* 18, 172-192.

Geary, R.G. (1958), "A Note on Comparisons of Exchange Rates and Purchasing Power between Countries", *Journal of the Royal Statistical Society Series A* 121, 97-99.

Gini, C. (1924), "Quelques considérations au sujet de la construction des nombres indices des prix et des questions analogues", *Metron* 4:1, 3-162.

Hill, R.J. (1997), "A Taxonomy of Multilateral Methods for Making International Comparisons of Prices and Quantities", *Review of Income and Wealth* 43(1), 49-69.

Hill, R.J. (1999a), "Comparing Price Levels across Countries Using Minimum Spanning Trees", *The Review of Economics and Statistics* 81, 135-142.

Hill, R.J. (1999b), "International Comparisons using Spanning Trees", pp. 109-120 in *International and Interarea Comparisons of Income, Output and Prices*, A. Heston and R.E. Lipsey (eds.), Studies in Income and Wealth Volume 61, NBER, Chicago: The University of Chicago Press.

Hill, R.J. (2004), "Constructing Price Indexes Across Space and Time: The Case of the European Union", *American Economic Review* 94, 1379-1410.

Hill, R.J. (2009), "Comparing Per Capita Income Levels Across Countries Using Spanning Trees: Robustness, Prior Restrictions, Hybrids and Hierarchies", pp. 217-244 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S. Prasada Rao (ed.), Cheltenham UK: Edward Elgar.

Hill, R.J. and M.P. Timmer (2006), "Standard Errors as Weights in Multilateral Price Indexes", *Journal of Business and Economic Statistics* 24:3, 366-377.

Iklé, D.M. (1972), "A New Approach to the Index Number Problem", *The Quarterly Journal of Economics* 86:2, 188-211.

Khamis, S.H. (1972), "A New System of Index Numbers for National and International Purposes", *Journal of the Royal Statistical Society Series A* 135, 96-121.

Marris, R. (1984), "Comparing the Incomes of Nations: A Critique of the International Comparison Project", *Journal of Economic Literature* 22:1, 40-57.

Neary, J.P. (2004), "Rationalizing the Penn World Tables: True Multilateral Indices for International Comparisons of Real Income", *The American Economic Review* 94, 1411-1428.

Schreyer, P. (2009), "Towards Measuring the Output of Health and Education Services", Working Paper, July 28, 2009 version, Statistics Directorate: Paris: OECD.

Sergeev, S. (2009a), "Aggregation Methods Based on Structural International Prices", pp. 274-297 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S. Prasada Rao (ed.), Cheltenham UK: Edward Elgar.

Sergeev, S. (2009b), "The Evaluation of the Approaches Used for the Linking of the Regions in the ICP 2005", unpublished paper, Statistics Austria, December.

Summers, R. (1973), "International Comparisons with Incomplete Data", *Review of Income and Wealth* 29:1, 1-16.

Szulc, B. (1964), "Indices for Multiregional Comparisons", *Przeglad Statystyczny* 3, 239-254.

Wang, C., S. Basu and J.G. Fernald (2009), "A General Equilibrium Asset Pricing Approach to the Measurement of Nominal and Real Bank Output", pp. 273-328 in *Price Index Concepts and Measurement*, Studies in Income and Wealth 70, Chicago: The University of Chicago Press.

World Bank (2007), *ICP 2003-2006 Handbook*, Washington D.C.: The World Bank.

World Bank (2008), *2005 International Comparison Program: Tables of Final Results*, preliminary draft, Washington D.C., February. http://siteresources.worldbank.org/ICPINT/Resources/ICP_final-results.pdf