

# Linking Country Groups in International Real Product and Purchasing Power Comparisons

Alan Heston and Bettina Aten<sup>1</sup>  
(May 29, 2002)

Introduction .....	1
Part I Are there Lessons to be Learned from Past Linking Exercises? .....	3
A. The 1980 Experience .....	3
B. The 1985 benchmark.....	4
C. Subsequent Benchmarks .....	5
D. Are There Guidelines from Past Experience? .....	5
Part II What is the Role for Spatial Chaining .....	6
Some Background .....	7
Why Not Link Multilaterally? .....	8
The Spanning Tree Method of Binary Linking.....	9
Some Anomalies of Spanning Tree Results.....	11
Choosing how to Chain or Group Countries.....	14
Regional Homogeneity.....	15
Part III Still Another Aggregation Approach to Linking .....	19
Why a Stochastic Approach?.....	19
The Results.....	21
Disaggregation.....	22
Conclusion .....	23
Annex on Linking within the Basic Heading Level.....	25
The EU Method of Estimating Heading Parities:.....	26
Implications for other Estimation Methods .....	27
How do we model the missing prices in estimating heading parities?.....	28
References .....	29

*“Finally the Chain Method is statistically laborious and inconvenient to apply in practice---so much so that it has been very seldom employed, in spite of the many years which have passed by since it was first recommended and the general approval (more than it deserves in my opinion) which it has received from theorists.” John Maynard Keynes (A Treatise on Money, 1930, p. 119)*

## Introduction

The above quotation from Keynes is primarily in the context of temporal chaining. However, it is in a chapter on *Comparisons of Purchasing Power* and his remarks on chaining would equally have been applied to spatial comparisons, which he also discussed.<sup>2</sup> Whether one accepts Keynes’s views, it is worth

<sup>1</sup> The University of Pennsylvania and Bridgewater State College. The authors would like to thank Robert Hill and Serguei Sergueev for comments on an earlier draft and Sergueev for sharing with us his paper on similarity indexes that we had not previously seen.

<sup>2</sup> This point is developed further in the Annex

reminding ourselves that many of the issues of concern in this and other papers have been around for some time.

Usually international comparisons are carried out at the level of GDP, at summary category levels of expenditure or production, or at the detailed heading level. Linking of groups of countries was explored in the 1975 benchmark comparison of 34 countries involving clustering techniques (Kravis, Heston and Summers, 1982). It became a more operational issue in the 1980 comparisons, which involved 60 countries. This was because there had been regional comparisons carried out by groups of countries like the EU, the OECD, Africa and ESCAP and these comparisons had to be bridged if there was to be a world comparison. Similar issues arose in the 1985 comparisons where fairly weak links had to be used to join various country groups in the world. It is in these comparisons that the Economic Commission for Europe began dealing with the so-called group 1 and group 2 countries, and the multiple answers one could come up with depending how Austria was treated. The OECD has grappled with this question since then both in integrating the EU country comparisons with the OECD countries, as well as more recently with the group 2 countries in Europe and the CIS countries in their 1990-99 benchmarks. (OECD, 2001a)

Problems that arise in linking countries and country groups at the level of GDP or lower expenditure groups, also arise at the more basic level of items used in price comparisons within basic headings. These issues are very important in improving the quality of purchasing power comparisons. The convention in international comparisons has been to treat these problems under item selection; so the linking of items will not be the focus of the paper. In order that this paper may be self-contained with respect to linking issues, comparisons within basic headings are treated in an Annex. The Annex suggests that chaining at the item level should also be considered within regions.

We summarize past efforts at linking in benchmark comparisons in Part I. However, the discussion does not take into account the EKS and G-K methods, except for comparisons, as these methods and results of applying them have been widely discussed. In the remainder of the paper we deal with two methods, spatial chaining and stochastic methods for linking country groups. In the past few years there has been a renewed interest in using binary links between countries to build up multilateral comparisons. These have been stimulated in good part by Robert Hill's proposed minimal spanning tree approach. One reason for this increasing interest is that more and more countries have taken to chaining their temporal price indexes and real product accounts (Hill, 1999a and b). Another reason is that as the number of countries participating in benchmark comparisons grows, concern about the reliability of the various multilateral methods of aggregation has grown. This has led to increased interest in spatial chaining methods with parallel efforts to find more satisfactory multilateral methods of aggregation. These issues are discussed in Part II.

The emphasis in Part III is the proposal of Prasada Rao that had its origins in his first work on international comparisons at the Indian Statistical Institute in Calcutta in the 1960s. Rao has proposed stochastic estimation of purchasing power parities (PPPs) across countries expanding on the country-product-dummy (CPD) technique that Robert Summers (1973) developed for Phase I of the ICP. More recently there has been interest in hedonic estimates of temporal price indexes such as Feenstra and Reinsdorf (2001). This interest is based on the concern with traditional index number techniques that implicitly assume that prices and quantities are observed without error. Once we recognize that substantial errors may be associated with the basic data used to construct indexes, it is natural to consider whether there are suitable stochastic models that could provide as good or better indexes than traditional index estimates.

## **Part I Are there Lessons to be Learned from Past Linking Exercises?**

Early work on international comparisons of real product began with binary comparisons between a pair of countries; if there were more than two countries, they were typically joined by binary comparisons involving a star or node country. However, as multilateral methods of comparison evolved they were given center stage during the first 25 years of the ICP. The 1970 and 1975 ICP benchmarks began with 10 countries and ended with 34 countries and had the advantage of having all prices and expenditures in one center, much like Luxembourg or Paris today for the EU and OECD. In the case of the 1970 comparison the main emphasis was to show that multilateral comparisons were feasible, that is the clear priority was to turn out a set of results. When the number of benchmark countries rose to 34 for 1975 it was possible to begin exploration of alternative ways of putting the world together. Later benchmark comparisons have never had the options available for estimating a world comparison based upon all prices and quantities available for all countries in one center for processing. Rather they have had to find ways to link the comparisons for various country groups.

### **A. The 1980 Experience**

The 1980 benchmark comparisons moved to a regional model with various UN Economic Commissions or groups like the EU and OECD making their own comparisons within a fairly common set of expenditure headings and items to be priced. However, at the end of the day, it was not clear if or how a world comparison might be developed for the 60 participating countries.

The method underlying the world comparison for 1980 (United Nations, 1985) relied on a set of price comparisons at the item level for a group of 20 so called core countries. There were a total of 60 countries in the benchmark comparisons, but it was a burden to try to obtain the data for the 50 countries for which detailed price data were, in principle, available. The decision was to obtain

detailed price data from enough countries in each region so that parities at the basic heading level would provide an overlap sufficient to develop world comparisons.

In the remainder of the paper the term bridge country will be used instead of core country. These bridge countries were typically judged to represent their region, e.g., East and West, and French and English speaking Africa. This usually provided 1-3 countries from each group, where EU and other-OECD are considered separate groups. This allowed direct estimation of parities at the detailed heading level for the 20 countries. The method of linking the parities to the remaining countries was done in a fairly mechanical manner. For example, if Senegal and Kenya were the African bridge countries, their heading parity would be expressed as a price level to the U.S, that is the heading parity divided by the exchange rate. The geometric mean of the price levels of Kenya and Senegal for that heading would then be the factor that would permit linking of the remaining African countries to the U.S. and the remaining regions.

The above procedure linked all 60 countries at the basic heading level, retaining their relative prices within their group for each heading. The aggregation was then carried out for all 60 countries using super-country weighted Geary-Khamis. This aggregation yielded total GDP for all the countries within each group. The regional results were then used to allocate the region's share of the GDP of the 60 countries, preserving the pattern from their regional study, a procedure termed fixity. Because these methods were developed after the fact, so to speak, there was an effort to spell out the next world comparison in advance.

### **B. The 1985 benchmark**

The framework of the 1985 comparisons was to try to establish high quality binary comparisons with selected countries in each region to a partner in another region. For example, Argentina and Germany and Kenya and the U.K., were two such proposed comparisons. Some of the links were in place as part of regional organizations such as Austria with Group II and the OECD and Japan with ESCAP and the OECD. The detail that each country grouping would use for its expenditures varied but the principle to be observed was that any additional headings created should be within the then existing world framework of about 150 headings.

Unhappily, many of the binary comparisons that were critical for linking regions were never completed. In addition, no South American countries were included in the 65 countries for 1985. The Caribbean countries joined the 1985 comparisons, but the binary comparison that would link these countries to Africa was never satisfactorily completed, so it was necessary to resort to use of a previous Jamaican benchmark to link this group.

### **C. Subsequent Benchmarks**

The OECD has successfully carried out benchmarks for their member and associated member countries and non-OECD countries covering a very significant portion of world GDP and over 50 countries in 1996, 41 countries in 1999, and several more in 2000. Their experience in this exercise should inform any future world efforts since the range of country incomes covered and wide range in statistical capabilities of the participating countries is representative of many of the problems at the world level. Also in putting together cross-sections in any years the OECD has examined several ways of bringing comparisons for other years to the reference year in order to include more countries.

The actual numbers illustrated in this paper are derived from an effort led by, but not embraced by, the World Bank to put together a world comparison of 115 countries for 1996. This patchwork benchmark involves only 32 headings, which more closely resembles the summary category level of most benchmark comparisons. It includes the OECD and Latin American countries for 1996, and updates of Africa, UNEC for Western Asia and East Asia and the Pacific from 1993 benchmarks. See Heston, Summers and Aten (2001) for more details.

### **D. Are There Guidelines from Past Experience?**

First, let us remember that in the future it is unlikely that we will have all prices and expenditures for all headings for all countries in one center. With current information-handling technology, there is no problem in handling such a volume of data and it should be a goal of the effort. However, it should not get in the way of carrying out comparisons with less than full information across regions.

Second, regional comparisons are desirable for their own sake. So any efforts to obtain world comparisons should be carried out as much as possible as a complementary activity to producing regional comparisons. Based on ICP experience of the last 25 years, the following is a short list of problems that have arisen in the implementation of previous attempts to build up world comparisons.

1. It is important to plan for a world comparison before, not after, regional benchmark comparisons are carried out.
2. For many reasons, there must be assurance that any linking between regions by pairing individual countries in each region, will happen.
3. Existing country links involve very large differences in economic structure and need to be supplemented by comparisons between countries more similar in economic level and structure. Both Austria and Japan have contributed resources to the linking efforts between the OECD (Austria) and other countries in Europe and to OECD

(Japan) with ESCAP countries. Because the economic level and structure of Austria and Japan are significantly above the levels of non-OECD countries in Europe and ESCAP respectively, they are not the ideal link countries. Developing other links, now possible with Korea, for example, can only improve the comparisons.

4. If linking is only done at the GDP level, they are very sensitive to which country is used to link regions. If more than one country is used for linking, then how multiple country links are to be combined deserves, at a minimum, some sensitivity analysis.
5. If linking is done at the detailed heading level, then it is still important to have countries whose economic structures bridges both regions. Linking at the detailed heading level, however, does have the advantage that it reduces the sensitivity of the aggregate results at the GDP level to the particular bridge country chosen.

Let us elaborate on point 4 above. Whatever producers of real product comparisons might say about the rich detail that underlies the benchmark comparisons, it is still true users most frequently want estimates at the GDP level. However, it does not follow that linking at the GDP level is necessarily desirable, even if that in the end is what is used. As noted in 5 above, GDP estimates will be less sensitive to the link country if carried out at the detailed heading level than at the GDP level.

Further, if linking can only be done at the GDP level, then consideration should be given as to what is the best way to do this. Clearly one can link through one or two countries in a region. But linking at the GDP level is basically seeking a scalar that converts the comparisons in one group of countries to a measure that is on roughly the same quantitative level as the other region. However, we know that the results are highly sensitive to which countries are used and how each is weighted. Given the sensitivity of the results it may be better to explore whether there might be a stochastic formulation that would allow more satisfactory estimates of the scalars needed to link groups of countries at just the GDP level.

## **Part II What is the Role for Spatial Chaining**

The adoption of chaining for price and quantity indexes over time has helped spur a resurgence of interest in spatial chaining. Despite strong support for spatial chaining, it is fair to say that the profession is not close to agreement yet on whether or how to do it. Temporal chaining has the advantage that one year follows the next, and this provides a natural, though not the only possible, sequence for linking over time.<sup>3</sup> Not so across space. There are natural

---

<sup>3</sup> A well-known problem with temporal chaining is that a disturbance in relative prices, like the recent oil price blip, may return to its original position after some time. However, a chain index will not return to its original position in its recording of the phenomenon. This may be less a

sequences by which to chain countries, like total population, total area, geographical contiguity, total GDP, or GDP per person. However, only the last would have any economic appeal. But it has a circular character since the whole process of spatial chaining is to arrive at a plausible cardinal order of the per capita GDPs of countries. So one needs the answer before beginning to answer the question.<sup>4</sup>

### **Some Background**

Because international comparisons were first carried out as binary exercises between two countries there was an initial move to make multilateral comparisons as a special type of spatial chaining, sometimes called the star method. The early work of Gilbert and Kravis (1954) used the United States as the center of a star involving the UK, France, Germany and Italy. Direct binary comparisons between the four European countries were not carried out, so the only way their relative per capita GDP could be derived was through binary comparisons with the United States. This use of one country as a node or center for a star has been employed in a number of other studies since that time.

For example, most member countries of the Council of Mutual Economic Assistance (CMEA) did very detailed price comparisons from the 1960s until 1990, though usually the results were not published. However, the method built up multilateral comparisons through binary comparisons of each member country with the USSR. The Economic Commission for Latin America (ECLA, 1967) carried out a study in the 1960s where the possibility of the United States playing a significant role was considered but rejected.<sup>5</sup> A star set of comparisons did take place with Austria as the center of Group 2 in Europe (Group 1 being the EU). During various benchmark comparisons, beginning with 1980, Austria was the center of a star that included Hungary, Poland, Finland, Bulgaria, and Romania. A characteristic of each of the star comparisons mentioned above is that the criteria for a country to be the node or center of a star was some combination of their willingness to do the work, or their political and/or economic centrality to the comparisons being carried out. From the standpoint of economic

---

problem for chaining over space.

<sup>4</sup> In an earlier paper (Kravis, Heston and Summers, 1978) KHS had used the ranking from the multilateral results for the 16 benchmark countries for 1973 as a basis for a chain binary and compared it with the results from a star system binary with the United States. The results were not encouraging in the sense that the resulting differences between the Paasche and Laspeyres indexes tended to be much larger than was the case for the direct star binaries with the United States. However, nodes or stars were not used in that chain, each country entering the chain twice except for the low and high countries that entered only once.

<sup>5</sup> For the period 1960-62 the Economic Commission for Latin America carried out a study of capital cities in 19 Latin American countries and in Houston and Los Angeles. In the end the study only carried out purchasing power comparisons excluding the United States cities. The method of weighting was to choose a representative basket for all 19 countries so that no country within the region played a larger role than another. (UNECLA, 1963)

similarity, comparing Hungary and Poland through Austria, or Italy and France through the United States, does not have great merit.

Another form of spatial linking has been between groups of countries. For example, Japan has been a link between ESCAP and OECD countries. Again, this has not been a very satisfactory method from an economic standpoint since Japan is at the high end of the ESCAP countries. This method was also used in the linking the 1985 results for Africa to the world, but here some of the links, like Pakistan and Kenya, were more plausible. Whatever the merit of past linking of regions, this is likely to be an increasingly important problem in future world comparisons.

### **Why Not Link Multilaterally?**

As multilateral comparisons have developed from binary beginnings it seemed logical to try to develop methods that dealt with the fact that binary comparisons between A/B, A/C, and B/C do not lead to a transitive result. That is the direct comparison of B/C will not generally equal the indirect comparison obtained by dividing A/C by A/B. Many investigations have been carried out on how to do this and many of the commonly used methods have been discussed by Irwin Diewert (1999) and Rao (2001). The broad results of all the methods support the most important finding of the ICP, namely that the price level (purchasing power divided by the exchange rate) of GDP systematically rises with per capita GDP, sometimes referred to as the Balassa-Samuelson effect (Heston, Summers and Nuxoll, 1994).

However, there are differences between the methods with respect to the total disparity of incomes between rich and poor countries, it being typically less with the G-K method using total country GDP as weights, and larger for the commonly used EKS method.<sup>6</sup> However, the differences between the two methods is much less than is the difference between either method and the result one would get using exchange rates (Kravis, Heston and Summers, 1982). Another issue between methods is how the sub-components of GDP are calculated; in G-K they are a by-product of the method while for EKS they have to be estimated directly by EKS, or indirectly.<sup>7</sup>

Another issue with multilateral methods is that as the number of countries in a comparison increases or decreases, the estimates for individual countries

---

<sup>6</sup> If the G-K method is used with the same weighting system as EKS, for example equal country weighting, then the differences between the two methods is slight. Put another way, the average prices that are used in G-K, are much closer to those of low-income countries when each country is assigned equal weight.

<sup>7</sup> When sub-components are directly estimated by EKS then the sum of the sub-components may be less or more than GDP as estimated by EKS over GDP, and the discrepancy may be 10% or more. It is also possible to distribute the EKS total by shares of sub-components at national prices but this means that the sub-components are not strictly comparable across countries. More discussion of this point is presented in Part 2.

may change. Since the number of countries usually differs from one ICP regional benchmark to another this does introduce potential variations in results across benchmark years that is regrettable and may be less than transparent to ordinary users of the results. When spatial chaining is used additional countries may be linked in without affecting the results of other countries, at least as a first approximation.

There is still a more fundamental difficulty in comparing countries namely holding the quality of goods and services constant between countries at very different income levels. One important influence on the quality of data, and its comparability across countries in all benchmarks, is the fact that the number of products and services and their quality are changing more rapidly than in the past. As is discussed in Heravi, Heston and Silver (2001) this has typically been handled by choosing items for comparisons that are of somewhat dated technology because they will be the common denominator across the countries. This is not obviously the appropriate choice, a point discussed further in the Annex.

### **The Spanning Tree Method of Binary Linking**

Robert Hill has spurred a revival of interest in spatial chaining, particularly his minimum spanning tree approach on which he has done considerable experimentation. This approach to chaining involves several dimensions (Hill, 1999a and 1999b) and results in a path by which to chain over space that involves the minimum number of binary comparisons to link all countries in a comparison. There are two elements in obtaining this chain. One is a minimization procedure that is used in conjunction with the second element, a plausible measure of economic similarity for each pair of countries. In Hill's spanning tree approach this second element is the Paasche-Laspeyres Spread (PLS) between each possible pair of countries. In an earlier paper Heston, Summers and Aten (2001) have looked at chaining using price similarity indexes as an alternative to the PLS. Sergueev (2001) and Diewert (2002) have reviewed a number of possible similarity measures that might be used, and have suggested alternative definitions that they consider preferable to the measure used in this paper.

In the Hill chaining procedure one country may also become a star for several other countries. However, there are two differences between this type of star country and those described above. First, in the spanning tree version there is a similarity in economic structure of the countries linked to a star country, whereas that has usually not been characteristic of links in the ICP in the past. Second, in the past almost all star countries were linked to countries within the same geographical area but this need not be the case with the minimum spanning tree chain.

An advantage of building up comparisons from a chain is that it may better control for quality differences between countries at different income levels. 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 being compared than across more economically diverse countries.

What criteria should be used to identify economically similar countries? As mentioned above, Hill has used the ratio of the Paasche to the Laspeyres index to obtain the spread (PLS) for prices or quantities. He in fact uses the logarithm of the absolute value of the spread so that it does not matter which country is in the numerator. His spanning tree approach minimizes the sum of the spread subject to the constraints that all countries must be in the chain, that no country can appear more than once, and that there are no closed or unconnected country loops. As noted in the Annex, Keynes thought that comparisons of two countries or two times would be more clear-cut the greater the overlap of goods purchased in the two countries or time periods. This might be another criterion for economic similarity.

For the purpose of this paper, the merits and shortfalls of chaining methods can be usefully discussed using results obtained using our definition of similarity or the PLS as criteria. Our price similarity index is defined as,

$$S_p = \frac{\sum_{i=1}^m w_i pp_{ij} pp_{ik}}{\sqrt{\sum_{i=1}^m w_i pp_{ij}^2 \sum_{i=1}^m w_i pp_{ik}^2}}$$

where  $S^p$  is the price similarity index between country j and k, and  $pp_{ij}$  is the price-parity in price level form for the i'th heading in country j, expressed as a deviation from the average value for the heading.<sup>8</sup> The weight for each of the m headings,  $w_i$ , is defined as:

---

<sup>8</sup> We are indebted to Erwin Diewert and Jim Cuthbert for pointing out that in our earlier work, the definition of the similarity index was not base-country invariant. In that version, the parity for each heading was expressed as the national currency units per US dollar, with the entry for the US for each heading being 1.0. However, because changing base country does affect the spanning tree path and the resulting multilateral comparisons, it is important to go with a definition of the similarity index that is base country invariant. In the present version each country's price level for a heading is expressed relative to that of the simple average of all the countries. As discussed in the text, there are alternative measures of price, and or price and quantity similarity that might be used as criteria for spanning trees, such as in Sergueev (2001).

$$w_i = \frac{(\text{exp}_{ij} + \text{exp}_{ik})}{2}$$

where  $\text{exp}$  is a country's heading expenditure share.<sup>9</sup> For each pair of countries, the average of their expenditure shares is used as the heading weight so there will be as many  $w$ s as there are country pairs. In the case of the price similarity index the spanning tree method can be applied where the sum of the price similarity indexes is maximized.

The character of the spatial chain that emerges from the spanning tree approach typically has a few countries that serve as nodes for a number of other countries. In effect there are several stars, some larger than others, as well as a number of countries that are at the extremities of the branches. And there are some countries that play a crucial role in the linking of several stars or clusters of countries. An essential point to keep in mind is that each country must be linked to at least one other country, but there is no limit (short of that set by the total country sample) to the number of countries to which a node may be linked. In practice, however, a node country will not have more than 4-6 links out of a total of 115 countries. In addition, nodes may cluster so that one can see if geography emerges as an important feature of the chain.

### **Some Anomalies of Spanning Tree Results**

Robert Hill and others have noted that the spanning tree approach can produce results that are not totally satisfactory. For example, Hill has found that the particular chain used to link countries is not robust across benchmarks. This may not be surprising, but it would be much nicer if the results were stable over time. Hill has suggested that the spanning tree could be implemented as a resource-saving device if one knew in advance the binary comparisons that were required. Thus, if Cambodia were only linked to Thailand, there would be no need for Cambodia to collect prices for items other than those relevant to Thailand. In the Annex it is suggested that a chaining of this type might still be appropriate at the item level, quite independent of the method of aggregation above the basic heading level.

Using either the PLS or price similarity criteria there are other features of the results that are problematic. What is troublesome in the procedure is that

---

<sup>9</sup> This definition of the weight is different from the one used in KHS (1982). Previously we had used the world average real expenditure share for each heading that came from total world expenditures from a G-K aggregation. We believe now that the average shares of each pair of countries at their national prices are a more appropriate weight. Also in KHS (1982) we did not use the heading parities, but rather the heading parities divided by the overall parity of the country so that it was a relative price that depended upon an overall PPP. We did this to provide an easy way for the US, the numeraire, to also have a similarity index with each country. And each country would have a similarity index with the world average, or Earthea, as we termed the world price structure. This was not done in the present paper, however. As noted in the previous footnote, some earlier versions of our similarity index were not base-country invariant.

often countries can take a pivotal position in the estimation, even when it is known that their database is not strong. In his work for the 1996 OECD benchmark countries for example, Hill (2001, p.14) found using the PLS that several countries had three links but that the country with the most links, six, was Greece. On many criteria, including overlap of items with other countries, Greece is not obviously the country to play a central role. In our own work for 1996 and described in Heston, Summers and Aten (2001), there were 32 heading parities and expenditure shares that had been put together by the World Bank for 115 countries from various regional comparisons that took place between 1993 and 1996. There are 6555 ( $115 \times 114 / 2$ ) possible links for each chain, that is, half the matrix of possible binaries. The first 114 links to join all countries without creating a closed loop form the minimum spanning tree or chain based on the PLS and price similarity indexes. The algorithm required going to the 2114<sup>th</sup> ranked pair (out of the 6555 possible pairs) for the PLS index and to the 2208<sup>th</sup> pair for the price similarity index.<sup>10</sup>

One way of judging a chain is by looking at which countries are linked frequently to others. For the PLS, there is some good news and bad news. The good news about Albania is that its links are all geographically close; the bad news is that it has four links so, like Uzbekistan with five links, these 2 countries are important in the chain in Europe and the former Soviet Union. However the quality of the statistical base of Albania and Uzbekistan is not as strong as countries like Hungary, which only has two links. The same problem exists in other world areas. Spain in the EU has the most links: five, but data quality in Spain and Greece, with four links, is probably lower than the Netherlands, which has only one link. Similar anomalies exist in Africa, where Tanzania has six links, the most of any country, and it in turn links Africa to Asia and the Middle East. Unfortunately the situation is hardly any better using price similarity indexes as the criteria for chaining. Chile and Spain both have seven links, while Bolivia, Ecuador, Peru and Ukraine each have five links in the spanning tree. And with the exception of Ukraine, the links are widely spread across the world regions.

Another aspect of the spanning tree approach is that it cannot make the claim that it only links one country with another because they are highest in price similarity or lowest in PLS. Any pair of countries will probably have considerable similarity but the following situation may well emerge. The chain begins with the highest link, say countries A and B. Suppose B is then linked to country E, which would in turn be most closely related to A. Since the loop of A, B and E cannot be closed, then E would be linked to the next best country, say D. It is possible that the highest link for country D would have been with country A, but this again

---

<sup>10</sup> In the PLS tree, the pair with the lowest Paasche-Laysperes spread is Peru and Ecuador, while the highest spread is between Georgia and Barbados, the latter being likely candidates to have very different price and quantity relationships. Using the similarity index tree, the highest and lowest pairs in the tree are Spain and Greece and Belize and the Bahamas. The latter is somewhat surprising given that these countries are from the same region.

would close a loop. So country D may not end up being linked with the country with which it is most similar.

The above are some of the less attractive attributes of chaining using the spanning tree technique on two candidate measures of economic similarity. Having said this, we will make clear below that there are several practical ways to deal with some of these problems that reduce the significance of these limitations on chaining. First, highly improbable links are reduced if chaining is done within regions first and then between regions. And as shown below, this substantially reduces the PLS or lack of similarity to be handled by chaining techniques. Similarly one can constrain the spanning tree approach to reduce the number of links in which a particular country is involved; thus countries with weak statistical systems might be constrained to enter into the spanning tree only once.

Having indicated some reservations about spatial chaining, it should be pointed out that the actual results of chaining are quite plausible. These results are presented in Table 2 in price level form, that is PPP/Exchange rate. and the actual chains are presented in Heston, Summers and Aten (2001). For the 115 countries in the empirical application the spanning tree approach provides a chain that generates a cardinal set of per capita income estimates for each country relative to the United States.<sup>11</sup> As a basis for judging how sensibly the chains performed they were compared with three alternatives, namely (1) Fisher indexes with the U.S. as the only node or star country; (2) EKS; (3) super-country-weighted G-K. A well-known result is that EKS tends to give lower estimates than G-K for low income countries. Using either the PLS or price similarity criteria, the spatial chains tended to be closer to EKS than to G-K for low-income countries, with a number of individual variations. Overall the chaining results were plausible, and displayed a number of characteristics similar to the multilateral methods.

We conclude that though there are limitations on the present state of the art with respect to chaining over space, the results are encouraging. This leads us to consider in Part 2 whether there are ways that spatial chaining might be improved and whether it is better in some types of applications versus others. Hill found that between the 1980 and 1985 ICP benchmark comparisons, the spatial chain was not stable. Note this does not mean that the ranking of countries by per capita income would necessarily be different using a chain from 1980 and one from 1985, but only that the actual links between countries that would be used to obtain the result would differ between the two benchmarks. It would be a tidier world if the links did not change between benchmarks, but this does not strike us as a major limitation of chaining.

---

<sup>11</sup> In the empirical work the actual measure of income was for technical reasons per capita domestic absorption.

## Choosing how to Chain or Group Countries

We will continue to leave the decision as to criteria for chaining open, but also continue to provide illustrations based upon the PLS or price similarity criteria. The minimum spanning tree is one way of using these criteria to arrive at a chain based upon binary comparisons of pairs of countries. One could also think in terms of a constrained maximization procedure in the case of similarity indexes. The spanning tree could be chosen so as to maximize the sum of the similarity indexes across the countries subject to a constraint that the minimum similarity index used for a country be less than some value or at its maximum. The object of such a constrained maximum would be to try to bolster potentially weak links in the chain. The PLS could be treated in a similar manner.

There are other techniques that might also be employed in identifying important country groupings. In Kravis, Heston and Summers (1982, Ch. 4), the PLS and price similarity were both used in cluster analysis to achieve groupings of countries, but not an explicit chain.<sup>12</sup> In this exercise the results of using cluster analysis were compared with those using a strictly geographical criterion or no grouping at all. These results are discussed in some detail because they are suggestive of still other ways that chaining might be carried out in the future.

For the 561 possible pairs of countries in the 1975 benchmark comparison the average value of the price similarity index was .754 with a standard deviation of .096. If one divided the 34 countries into 6 geographical regions, there would be 165 pairs of countries, with an average price similarity of .803 and a standard deviation of .082. Using cluster analysis yielded 6 groups with 134 pairs where the average price similarity index was .840 and the standard deviation .050. Clearly cluster analysis yields groups of more similar countries and also improves upon a strict geographical division, though less so for the PLS.<sup>13</sup> That is the cluster analysis increases the average price similarity index from .754 to .840, or .086 while the straight geographical coverage raises average similarity from .754 to .803, or .049. So a regional grouping by itself accounts for 57% of the

---

<sup>12</sup> It should be pointed out that the PLS may be expressed in the way we have described it up to now, namely a number that takes on larger values for pairs of countries that are less similar economically. In KHS (1982, Ch. 7) the measure of the PLS was the ratio of the per capita income of the lower income country to that of the higher income country; consequently, more similar pairs of countries took on a value of the PLS closer to 1, and less similar pairs of countries had lower values. The advantage of this formulation versus the absolute value of the log of the ratio used by Hill, is that it was parallel in direction with the price similarity index. For the 34 countries in KHS (p.109) the regression of the PLS on the price similarity index across the 561 pairs of countries was  $0.45 + .37 * \text{Price Similarity}$  with  $R^2 = .47$ .

<sup>13</sup> A similar exercise was carried out for the PLS and the pattern was similar. However, whereas a straight geographical division of countries accounted for 57% of the reduction in the average price similarity index as discussed above, for the PLS the same division into regions accounted for 87% of the reduction in the PLS. (KHS, p. 109) To anticipate later discussion, the average values of the similarity indexes (PLS) for the 1975 benchmark were lower (higher) than the exercise we have reported for 1996. This is most likely due to the fact that the 1975 benchmark involved over 150 common detailed headings, and the 1996 exercise only 32 headings.

increase in the average similarity index (.049/.086 \* 100) achieved by cluster analysis.

This discussion points up the potential power of techniques like principal components to sort out some of the relationships of economic similarity between countries. While these techniques do not themselves provide an explicit chain, they do offer breakdowns of the countries into groups where chains could be developed. However, a major practical limitation on the use of such methods, including the minimum spanning tree approach, is that the organization of the benchmark work may substantially limit the options that can actually be implemented. This may be illustrated with a case that turned up in the 1975 benchmark. One fairly important link that emerged from cluster analysis based upon the PLS was between Uruguay and Iran. Even if such a result between Iran and Uruguay were replicated in other studies, it is unlikely that either country or any international organization would see much merit in these countries engaging in an intensive binary comparison.

### **Regional Homogeneity**

One great advantage of multilateral methods is that the actual estimation procedure does not require discretion on the part of those producing the results. There are still a great many judgements made in the process of preparing the inputs, including item and outlet choice, the handling of non-priced services and the like, but once that is done, the computations can be mechanically carried out. However, while this has been true for benchmark comparisons within regional groups like ESCAP and other groups like the OECD, the joining of comparisons across groups has always required a number of subjective judgments. The likely world in which future comparisons are carried out is one where they are initiated at a regional or country group level. And it is unlikely that a regional group will want an intermediary country from another group playing a role in their comparisons, which was the type of chaining model discussed in Part II.

These country groups might be thought fairly homogeneous, but typically that is not the situation. The case we have made for spatial chaining to overcome problems of holding constant the quality and range of goods compared apply within most country groups. The OECD core countries are fairly diverse across space and economic structure. The CIS and Eastern European countries have also undertaken comparisons in collaboration with the OECD and this greatly increases the heterogeneity of the total group of 50 countries. ESCAP includes Hong Kong, Singapore and Japan at the top and Nepal, Bhutan and Laos at the low-income end. The Caribbean ranges from Haiti to Bermuda and Western Asia is equally diverse.

Our conclusion is that generating a chain across all countries of the world in a benchmark is not a likely nor necessarily desirable outcome. It is not likely given the way that ICP benchmark comparisons are organized. And it may not

be desirable because existing algorithms to generate chains produce results that are only marginally better than chains built up from the regions. This last point is based upon some of the evidence in Table 1 below from the 1996 comparisons.

<b>Table 1: Spanning Trees versus Regions</b>				
	Similarity Index Averages by Groups			
	All Pairs	Tree	Regional Groupings 15	7
# Pairs	6551	114	744	1044
mean	0.9251	0.9939	0.9606	0.9580
st.dev	0.0668	0.0060	0.0469	0.0452
cv	7.2%	0.6%	4.9%	4.7%
	PLS Averages by Groups			
	All Pairs	Tree	Regional Groupings 15	7
mean	1.8677	1.0445	1.2836	1.1947
st.dev	0.9122	0.0667	0.4021	0.3135
cv	48.8%	6.4%	31.3%	26.2%

In Table 1 we have examined measures of the average price similarity and the average PLS for various combinations of pairs of countries. First we consider in column (2) of Table 1 all pairs of countries, and in column (3) the minimum spanning tree of 114 countries. In columns (4) and (5) the pairs involved with 15 regional groups based solely on geography, and finally a set of 7 groups of countries that have in the past actually implemented regional comparisons.

The results in Table 1 show that the minimum spanning tree increases the average similarity index and greatly reduces the average PLS compared with what is achieved over all possible pairs of countries. As EKS uses all possible pairs of countries, this does provide some indication of why using this technique over all countries may introduce noise in the quest for transitivity.

The main point that we wish to make about Table 1 however, is the degree to which breaking the countries into regions substantially reduces the average PLS and increases the average similarity index. Consider first the PLS. Using all pairs the average is 1.8677 which is reduced to 1.0445 when the minimum spanning tree is used, a reduction of .8232. However, most of this reduction can be achieved by moving to regional groups. Thus moving to 15 geographic groups reduces the average PLS by .5841 or by 71% of .8232. When we go to the 7 implementing groups, it accounts for 82% (.673/.8232\*100) of the reduction in the PLS achieved by the minimum spanning tree.<sup>14</sup> But this

<sup>14</sup> The fact that division of the countries into the 7 groups does more to reduce the average PLS than the division into 15 groups is a comment on the degree to which implementing groups of countries tend to be more homogeneous in putting together the basic data for the benchmark comparisons. However, this pattern does not hold for the price similarity indexes in Table 1.

only tells a part of the story. Suppose that within the geographical regions a minimum spanning tree were constructed. We have not done this for all 15 regions, but it will be enough to illustrate with one region, East Asia where there are only 4 countries. The average PLS using the minimum path is 1.2772, while using all 10 pairs in the region it is 1.4422. So that if one chose to work with pairs of countries within regions one would reduce much of the PLS and going to the minimum tree within the region would reduce it still more. Put another way, these results suggest that not much is lost by applying the minimum spanning tree technique only within regions, rather than across all countries.

For price similarity indexes the increase in the average is also improved by going to regions, though the percentage effect is somewhat smaller. That is, the improvement in Table 1 from all pairs to the minimum spanning tree is .0688. Of this .0355 or 52% of the increase in average similarity index is obtained by going to 15 regions and .0329 or 48% by going to 7 country groups. This is not as large as for the PLS but there is also less differences between the similarity indexes than the PLS.<sup>15</sup> As with the PLS there is a further reduction of the average similarity indexes if one were to apply the minimum spanning tree within regions. As an illustration for Africa and the Middle East, a group of 11 countries and 55 possible pairs, the minimum spanning tree with the region yields an average similarity index of 0.9803 versus 8929 for the 55 pairs. So as with the PLS much of the gain in economic similarity can be obtained from regional groupings of countries, and applying chaining within the country groups.

While multilateral techniques have been and can be applied at a regional level, binary chains within regions should seriously be examined as an alternative. First, there is substantial diversity within regions so it may be very useful to build up regional comparisons on the basis of binary comparisons of economically similar countries. This also might allow some countries with less diverse markets and modest statistical capabilities to carry out benchmark comparisons for fewer basic headings by linking to similar countries that are slightly higher on the economic and statistical ladder. The choice of such pairings could be based upon criteria like the PLS spread, though in many regions there may be consensus in a region on where to begin the chain.

Given the regional structure of benchmark comparisons, a goal in putting together world comparisons should be to reduce the need for discretionary judgements as much as possible. It is in this spirit that a proposal is offered below on how spatial chaining and regional or country group comparisons might be combined

The proposal is a variant on how the 1980 regional benchmark comparisons were to be linked to form a world comparison where the link

---

<sup>15</sup> For example the lowest similarity index observed is .398 with average as in Table 1 of .925. If the PLS were also constrained to be between 0 and 1 across all pairs, the lowest value (biggest spread) would be .306 and the average would be .530.

between regions would be based upon a set of bridge countries from each region. These countries would provide price data in as much detail as possible with respect to both market and item characteristics as described by Heston (2002). This would permit estimation of a hedonic regression either by the country or by a coordinating group for the items in a basic heading. The availability of such an estimating equation at the basic heading level would also allow comparison of estimated prices of items within each heading for countries in each region that had a strong enough database to permit such estimation. These item price estimates could in turn form the basis for a comparison of core commodities across the each region. Assuming the availability of core item prices that could link regions becomes a part of future benchmark comparisons, it would provide a basis for either binary or multilateral comparisons.

Use of a set of bridge countries from each region would permit collecting a common set of prices with enough overlap of items that parities can be obtained across regions at the basic heading level. These parities could be obtained at the basic heading level by either the CPD or EKS methods. This overlap of core countries provides a basis for linking regions permitting the expression of basic heading parities in a common currency unit, for illustration, say Euros. However, there remain decisions to be made as to exactly how this link will be made. For example, suppose there were 20 bridge countries that were willing to provide enough item price overlap so as to represent 5 regions comprising the world. The countries representing the regions will be from the region but will not necessarily represent them in being economically similar to typical countries in the regions. The following are three ways that the linking might be done. In the discussion below, suppose that for a typical region, there are four bridge countries 2 upper-income countries, 1 upper middle-income, and 1 lower middle-middle. This might be a likely distribution since few very low-income countries are likely to have the ability to supply enough prices to become a core country.

One method would be to average the relationship of the 4 bridge countries of the region. Suppose for a basic heading like fresh fish the EKS or CPD estimation gave the first 2 countries a price level of 125 with respect to the Euro, the third country 110, and the fourth country, 100. The simple average price level would be 115. The fish parities of the remaining countries in the region would be multiplied by 1.15, to express them relative to the Euro. Variations on this approach might take account of the population or total GDP of each of the bridge countries.

A second method would be to weight the price levels in each bridge country according to the proportion that their economic structure represented the region. This could be done by treating countries as the unit, by using total population or total GDP as the weight.

A third approach would be in terms of chaining. For example, if a spanning tree exercise had been carried out within a region, then each core country would have been paired with one or more countries that could be used to convert their binary comparison within the region into a relationship of both to the Euro. This could be done for each bridge country permitting a binary based linking to the world. A variation on this approach that we prefer is set out in the Annex where the chain would be determined by the percentage of overlapping products between two countries relative to all possible items within a detailed heading.

These methods should be agreed upon in advance so that as little discretion as possible is required to link regions. The third approach seems better in this regard. However, it does depend on the countries in a region having adopted some type of spanning tree approach to put together their within-group estimates.

### **Part III Still Another Aggregation Approach to Linking**

In this part of the paper we present some preliminary results using the method that Prasada Rao has proposed, a variation on the hedonic regression that Robert Summers originally suggested. The background for the CPD method is discussed in the Annex and elsewhere. Rao (2001) provides a framework for this method, as do Rao and Timmer (2001) who apply the method to production side estimates of real product and PPPs. The present application will be to the world aggregation for 1996 so that it will be simple to compare with chaining techniques and other aggregation methods. However, as an aggregation technique it could be applied in linking regional results at the detailed heading level or higher levels of aggregation.

What would recommend the type of approach Rao has advocated? First, Rao has shown that under certain circumstances the parameter estimates from this method are the same as a log form of the G-K method so it is not such a radical departure from other aggregation procedures. However, why would you want that in preference to existing methods like EKS or G-K, or a chain approach? Two points developed below are (1) that there are advantages to a stochastic approach when the underlying inputs to the indexes are observed with errors; and that (2) this method also provides a more natural approach to disaggregation than either the EKS or chain approach.

#### **Why a Stochastic Approach?**

This seminar is hearing presentations on comparison resistance services, construction, and other areas of price comparisons that are difficult to conceptualize, let alone carry out without substantial uncertainty about the underlying price ratios and heading parities. Other papers address estimation of

heading parities where it may be easier to identify the items, but still difficult to estimate because of price variation of items across outlets and space. With the errors surrounding the input parities, and the expenditures too for that matter, it is curious that we, present company very much included, continue to estimate index numbers without much attention to the error associated with them.

There has been a major research effort in recent years to use hedonic techniques in time-to-time price indexes that is beginning to carry over to spatial comparisons. Feenstra and Reinsdorf (2001) have argued for a stochastic approach that has at least as much theoretical justification as the EKS or similar binary based approach. Their current formulation requires an ordering of observations as in temporal chain indexes so that its direct application to spatial comparisons requires a spatial chain for implementation. It is not possible in this paper to further develop their formulation but it is an area for future research.

A major advantage of a stochastic formulation is that it provides a sense of the errors involved in the final indexes. It is an illusion of most index numbers that they are estimated without measurement error. Put another way in the axiomatic approach to index numbers such as Diewert (1999) the list of criteria for index numbers never includes a tradeoff between measurement error and other characteristics. Clearly an advantage of a stochastic formulation is that it provides some insight into the range of estimate for the PPPs that are the end product of the work.

The particular formulation suggested by Rao has an additional advantage associated with any simultaneous estimation of PPPs and heading coefficients, namely that it permits a natural disaggregation of the results. In the original CPD formulation of Summers, coefficients on heading parities and coefficients for each item within the basic heading were estimated. An analogous set of coefficients is generated by applying the method to a set of heading parities, though in the case of parities we also have an associated set of expenditure weights.

In the illustration below these expenditure weights are explicitly used in the CPD approach, as Rao (2001) and Rao and Timmer (2001) have done. Each country has the same importance in the estimation. Thus the weighting is similar to that in EKS, a chain comparison, or a G-K aggregation where each country has the same weight. The equation below is estimated across the 115 countries and 30 headings, where  $p_{ij}$  is the heading parity,  $\hat{a}_j$  is coefficient on each heading  $z$  and  $\hat{a}_i$  is the coefficient on each of the country dummies except the numeraire. The estimation of (1) minimized the weighted sum of the squared residuals where the weights are the expenditure shares.

$$(1) \ln p_{ij} = \sum_{i=1}^n b_i z_i + \sum_{j=2}^m a_j D_j + e_{ij}$$

One point to note about this application is that the input data are the same as used in standard aggregation procedures, namely parities and expenditures at the detailed heading level.

## **The Results**

An empirical point to be noted about the estimation of (1) is that the  $p_{ij}$ s are in price level form with the heading parities in national currencies divided by the exchange rate. In practice these vary from about .30 to 1.80. This reduces the variance to be explained as compared to PPPs, where the dependent variables are say, the Italian lira/\$ and UK pound/\$, and the variance across currency units is 1 to 2000. As a consequence, our estimating equation removes the correlation associated with country effects explaining currency units.

The illustration presented in Table 2 displays the results of the stochastic estimates in column (8), where the countries have been ranked by per capita domestic absorption as a percent of the U.S. The estimate of income uses the binary Fisher with the United States so as to be independent of any of the methods of chaining or aggregation. For each method the price level is given as a percent of the United States so that, reading across the columns, lower values are associated with higher income estimates and *vice versa*.

Two examples of binary chains are given in columns (6) and (7), the first based on the criterion of price similarity and the second on PL spread. In columns (3) the Fisher price level with the US is provided, the same price level associated with the income index in column (2). For reference the price level by the EKS method is given in column (4) and by the G-K method with super-country weights in column (5).

At the bottom of Table 2 some summary statistics are provided as simple averages of the 115 countries. When comparing all countries the CPD approach is somewhat below EKS and the Chain methods, an average price level 64% of the U.S. However, this is because it estimates the upper income countries with lower price levels (115% of the U.S. and hence, higher incomes) than any other method. For low-income countries it provides the same estimated price level (44% of the US) as Fisher, and a higher price level (so lower incomes) than EKS. So the spread in estimated incomes across countries is larger by the CPD approach than with Fisher or EKS, and much larger than by G-K.

The CPD using expenditure shares is a very simple form. What is the interpretation of these estimates compared to other methods? Looking first at the regression statistics, how do we interpret the unexplained variance that is

35%? That is, the country dummies and the expenditure shares of the headings account for 65% of the variance in the 3402 heading parities entering into the CPD. One potential advantage of the stochastic approach is that other variables might be added to the equation and their importance tested by measuring the reduction in residual error.

However, additional variables cannot be country variables, such as average rainfall in a country (an unlikely candidate), because it would be perfectly collinear with the country dummy. Introduction of other variables that differ by both heading and country, like the coverage measures of Rao and Timmer, would be a valuable area for future research. Sergueev has also suggested measures associated with each heading, like variability of observed price ratios across items, which also seems like a good candidate for investigation.

### **Disaggregation**

We show less caution than Rao and show in Table 3 how the CPD might be used at the heading level. In Table 3 we only illustrate it for C, I and G, but the approach could be done at any level of aggregation. First at the highest level of aggregation in this exercise, Domestic Absorption (DA), we accept the country PPP estimates from the CPD to convert national currency to international dollars. We then use the heading parities in the way one might use international prices, namely to value the national currency headings divided by the heading price levels. The sum of these estimates will not equal the converted DA level because the CPD coefficients are multiplicative. However, it is a simple matter to normalize these totals to provide a consistent set of numbers. While it is not strictly correct to treat them as additive, it is not a bad approximation. And it is very different from applying national currency shares to DA.

Table 3 provides the share of Consumption, Government and Investment in national currencies. It then uses these shares as a denominator to divide the shares that one obtains from the CPD estimates as described, and from the usual G-K estimates, this being a likely comparator. The ratios to national shares vary a great deal from country to country and some of the more outlandish observations are less a comment on methods than the data inputs from the countries. However, there are two points that we would make about Table 3.

First, the differences between the CPD and the G-K ratios for C, I and G are fairly close together and differ in a desirable way. That is, in the arithmetic G-K there has probably been an overstatement of real shares of investment in rich countries and an understatement for government, while the opposite has been the case for poor countries. The shares based on CPD somewhat reduce these tendencies.

Second, while there may be better ways to link countries and estimate shares than the CPD method, it would seem to us that EKS cannot make that claim. The CPD and EKS estimates at the total GDP level are very similar. And the CPD approach offers a fairly natural way to produce heading estimates. The use of the EKS method at each level of aggregation is deficient because it does not allow us to know the relative prices of the various aggregates of GDP. Thus we would conclude that even the simple CPD approach dominates EKS in terms of its desirable attributes.

This contentious comment aside, our conclusion is that the approach that Rao has advocated deserves much more investigation. It is certainly consistent in spirit with the movement to stochastic methods of estimation of price and other indexes over time. It is not theoretically pleasing, but there may be variations that can be developed along the lines that Feenstra and Reinsdorf have been exploring.

## **Conclusion**

The conclusions are at several levels. The first concerns the overlap of regions. Considerations from earlier studies strongly suggest the desirability of linking regions at a detailed level, rather than at the level of GDP. While there are a number of ways this might be done, so long as we have item price data on similar regions across a group of bridge countries with two or more countries from each region, a variety of linking approaches may be used. If matching of items is to be carried out, the specification should include an item description and type of outlet. If item data includes market and item characteristics, it may be possible to estimate heading parities using a hedonic approach.

Second, it seems impractical and probably unsatisfactory to use a chain technique across all countries in all regions. However, spatial chaining may play an important role at two levels. One would be as a way to make item price comparisons within basic headings, about which the Annex offers some comments. Another is to link in countries that may be marginal to a regional comparison in the sense that their data are not timely or their coverage of items and expenditure headings are limited.

Third, once we have a world of heading parities across countries in different regions, we are back to a number of alternatives for putting together regions. Our preference is to carry out an aggregation over the world in which each country has a resulting GDP and components. This would provide a basis for estimating the regional income of each group of countries organizing their own benchmark comparison. Regional fixity requirements may be respected by the region distributing their share according to the results of their own regional comparison.

The fourth conclusion coming from researchers is of course we need more research. We have suggested there are a number of roles for chaining, more limited than those that have been presented in Table 2, and these should be examined both below and above the basic heading level. Our experiments with the CPD-Rao approach to aggregation, suggest there are a number of possibilities for using stochastic methods in an area that has probably concentrated too much on index number formula. Another related area of research using stochastic techniques is suggested by the work of Rao and Timmer and their concern with how well item price comparisons represent a branch of industry or heading of expenditure. If we can readily quantify the error associated with different headings, then a stochastic approach would allow us to integrate these as weights in estimating parities. And finally, if we are to pursue a stochastic approach, then examining techniques that are less rigidly structured than CPD and have more basis in economic theory, while not necessary, would be highly desirable.

## Annex on Linking within the Basic Heading Level

This Annex is concerned with linking below the basic heading level and how this relates to the quality problem and the estimates of heading level parities. Consider the common problem of comparing price levels in a country between  $t-1$  and  $t$ . Schematically we set out below the essence of the problem as described by Keynes, who recommended that practitioners choose those parts of the expenditures in two positions that are identical, “supplementing this by a list of the expenditures discarded and added (so as to enable a general judgment to be made as to the extent of improved opportunities)” (1930, vol 1, p. 119).<sup>16</sup>

Group <i>D</i> (ropped)	Group <i>I</i> (dentical)	Group <i>N</i> (ew)
Goods in $t-1$ only	Goods in both $t-1$ and $t$	Goods available in $t$ only
_____ /	_____ /	_____ /

A dimension of chaining over time is that the  $t-1$  and  $t$  comparison is natural and probably provides the largest number of goods in Group *I* (in this discussion we will for convenience present our ideas as though the expenditure share is proportional to number of goods within a heading).

One can think of spatial comparisons as an analogous problem. Between any pair of countries would it be meaningful to choose a chain based on the size of group *I* to the total items in a detailed heading? Since the European Union (EU) has examined this problem in considerable detail it is useful to present their framework. The reader is warned that this exposition probably will not get the EU method quite right but for purposes of this discussion, the following schema should capture its essence. Consider countries A and B and all items for all countries including A and B in a comparison for a given basic heading.

	Goods Priced in A	Not Priced in A	
	Important	Not Important	
Goods Priced in B			
Important	$I_{A^*B^*}$	$I_{A/B^*}$	$D_A: N_{B^*}$
Not Important	$I_{A^*B}$	$I_{A/B}$	$D_A: N_B$
Not Priced in B	$D_B: N_{A^*}$	$D_B: N_A$	$D_{A/B}$

<sup>16</sup> For a recent discussion of this problem see (Pakes, 2002).

First, the starred (\*) items are those considered important in the consumption of a country within a basic heading. Other items may be priced and are commonly available but not important. Finally there are items that are available and not common, or that are simply unavailable; they are not priced by either country A or B. In terms of the temporal schema above, the goods in  $I$  are:  $I_{A^*/B^*} + I_{A/B^*} + I_{A^*/B} + I_{A/B}$ . The difficult problems arise in interpreting what are the equivalents in cross-country comparisons of  $D$  or  $N$  goods in temporal comparisons. Considering the items within any basic heading that are not compared between any pair of countries for a basic heading, there seem to be 3 principal explanations.

(1) Country A might not price an item that is priced in B either because (a) the relative price is so high in A that there are few purchases; or (b) the relative price in A is similar to B but A's income is so low they substitute other items. Clothes driers might be an example.

(2) The item is truly new or high tech, and even though the price might be similar in both countries, incomes are low in A and they choose not to consume the item in any quantity. And more likely there are complementary income and price effects operating to make purchases of high tech products more likely in rich than poor countries. Flat screen displays might be an example.

(3) There are cultural or taste factors including prohibitions or proscriptions of items like alcohol or pork; or simply styles of dress or cuisine. While most comparisons eschew items that are passing fads, some fashions have longer lives than others.

Clearly Keynes's observation is less relevant to the price and cultural effects in (1) and (3) above. If this is a useful framework to think about item selection then is there a significance to the proportion that  $I$  goods are to the total in a heading? To be more precise, and using the EU framework the goods entering into the estimation of the binary estimation between A and B, are:  $I_{A^*/B^*} + I_{A/B^*} + I_{A^*/B}$ . Let  $p$  express these as a percent of possible items compared in a heading.<sup>17</sup>

### **The EU Method of Estimating Heading Parities:**

The EU method consists of estimating parities between all possible pairs of countries for which  $p > 0$ . The EKS procedure is then used to produce one transitive set of parities for all the countries for that heading. One limitation of this procedure is that all the binaries for which  $p > 0$ , enter into the EKS estimation with equal importance, even though  $p$  may be very different among the binaries. This procedure could be improved by using  $p$  as a weight in the

---

<sup>17</sup> The number of items entering into  $I_{A/B^*}$  and  $I_{A^*/B}$  may be different: therefore the geometric mean for a heading for A/B need not be the same as for B/A, this being the reason for distinguishing between starred and non-starred items in the EU schema.

EKS or by a linking or chaining rule. For example, one might exclude all binaries from the EKS for which  $p$  is less than some number, for example, 50%.

Rao and Timmer (2000) have used two concepts of what they call the coverage ratio as applied to production data for 13 branches and 7 countries for 1987. The first,  $c_i$ , ratio uses the proportion of a country's expenditures covered by items entering into price comparisons. The other concept,  $c_{ij}$ , is the average ratio of output covered by the items matched between the two countries. It is this latter concept that one would want to use as a weight in any comparisons built up from pair-wise comparisons such as has been discussed for the EU. The coverage ratio,  $c_i$ , would be appropriate for strictly multilateral systems. As mentioned above, Sergueev has suggested that in addition to the coverage ratio, other criteria might be considered including the variance of price ratios within a heading.

In practice however, calculation of  $p$  is not as simple as implied in this discussion, especially in the EU. The reason for this is that many of the specifications are only slight variants on each other. An additional suggestion about this question is given below, namely that what is considered an item within a detailed heading needs rethinking, or needs to be spelled out more clearly than is typically done in benchmark comparisons.

Another aspect of the EU procedure is that items  $I_{A/B}$  do not enter into the binary comparisons, this being a form of weighting within basic headings. Two observations may be made here. First this information could be retained and given less importance, but not zero, in estimating the binary parity. And second, the method appears to assume that we observe prices without error, when we know that there are many departures of prices used from true prices. For this reason there is a cost to neglecting the  $I_{A/B}$  information if all of the prices are observed with similar sampling error; This is because the more observations bearing on the desired parity for the basic heading the less will be the error of the estimated parity for the basic heading, a point made that Sergueev has also brought out in correspondence.

### **Implications for other Estimation Methods**

In the CPD method of estimating heading parities as carried out in Phase I of the ICP benchmark, the weighting was the opposite of what is being recommended above, namely the importance each country was inversely related to  $p$ . The reason for this was that in the original CPD formulation the matrix of prices of  $n$  countries and  $m$  items had holes (the rationale for this was that coffee had been spilled on the full matrix) so that missing item prices were assumed to be randomly distributed. For this reason it was felt that each country should receive equal weight in the estimation, and this was achieved by weighting the importance of each item price for a country inversely to the number of prices within a heading.

Needless to say the argument presented here suggests that the original method of weighting CPD was modeled inappropriately.<sup>18</sup> Probably the number of prices provided for each heading by a country is the more appropriate weight. In this view the holes we observe in a heading of  $n$  countries and  $m$  items are generated by the 3 factors discussed above, a subject next addressed.

### **How do we model the missing prices in estimating heading parities?**

Probably the more meaningful way to think of the number of items within detailed heading is the ELI or Entry Level Item of the BLS. This formulation is consistent with the formulations of Michael Ward (2002) and Heston (2002) of how we should think of items. An item is the bundle of market characteristics like outlet and item characteristics like size, packaging and brand. With this view of the nature of an item, how should we think about  $p$  between countries?

Returning to the three reasons we discussed above for missing prices, matters like culture and taste appear to introduce no particular bias to estimation procedures. However, in the matter of high tech or new products, the probable direction of bias is to understate the parities of lower income countries because it is likely that these items are excluded from comparisons even when they have acquired some importance in purchases. This will occur because of the usual lags in the introduction of new goods into temporal price indexes. If such items were to enter spatial comparisons, it is likely it would lead to a higher estimated parity for the heading than usually obtained. With respect to factor (1) above, if there is a direction of bias it is probably the same as for new goods, but how one would measure it is unclear.

The conclusion would seem to be that in estimating basic heading parities consideration should be given to methods that take account of  $p$  in their formulation. In this the EU procedure has taken important steps but it too could be improved in implementation. The CPD procedure that provides weight according to the number of items priced by a country seems the right direction when using this technique. Better still would be to rethink what we mean by an item within a basic heading and to systematically consider  $p$  in how we estimate heading parities. This Annex is intended as a step in that direction.

---

<sup>18</sup> The possibility that the holes in the price matrix had both a systematic and random component was recognized in the early phases of the ICP. Discussions took place and some experiments were carried out and still others were considered. However there were pressures to get the comparisons completed; so having squarely faced the problem, we walked away from it.

## References

Diewert, W. Erwin (1999, "Axiomatic and Economic Approaches to International Comparisons," in International and Interarea Comparisons of Income, Output and Prices, A. Heston and R. E. Lipsey editors, NBER and CRIW, Studies in Income and Wealth Vol. 61. University of Chicago Press.

Diewert, W. Erwin.(2002) " Similarity Indexes and Criteria for Spatial Linking" International Conference and Expert Group Meetings on the International Comparisons Program, World Bank, March 11-15.

Gilbert, Milton and Irving B. Kravis (1954), An International Comparison of National Products and the Purchasing Power of Currencies: A Study of the United States, The United Kingdom, France, Germany and Italy, OEEC, Paris.

Heravi, Saeed, Alan Heston and Mick Silver "Using Scanner Data to estimate Country Price Parities," World Bank-OECD Seminar on Purchasing Power Parities, Washington, D.C. [www.oecd.org/std/ppp/MTG2001/index.htm](http://www.oecd.org/std/ppp/MTG2001/index.htm), January 30 – February 2, 2001.

Heston, A. (2002) "What Improvements Can be Made Quality and Usefulness of Prices Collected for Commodities and Priced Services for PPP Estimation" International Conference and Expert Group Meetings on the International Comparisons Program, World Bank, March 11-15.

Heston, A., R. Summers and D.Nuxoll (1994) "The Differential Productivity Hypothesis and Purchasing Power Parities: Some New Evidence," Review of International Economics, October.

Heston, A., Robert Summers and Bettina Aten (2001), "Stability of Country Price Structures: Implications for Spatial-Temporal Comparisons," UN Economic Commission for Europe Statistical Reporter, May. Also on OECD website.

Hill, Robert (1999a), "Comparing Price Levels Across Countries Using Minimum Spanning Trees," Review of Economics and Statistics, 81, No. 1.

\_\_\_\_\_ (1999b), "International Comparisons Using Spanning Trees," in International and Interarea Comparisons of Income, Output, and Prices, ed. A. Heston and R. Lipsey, Studies in Income and Wealth, vol 61, NBER, Univ of Chicago Press.

\_\_\_\_\_ (2001), "Linking Countries and Regions using Chaining Methods and Spanning Trees," Joint World Bank-OECD Seminar on Purchasing Power

Applications, Washington D.C., [www.oecd.org/std/ppp/MTG2001/index.htm](http://www.oecd.org/std/ppp/MTG2001/index.htm), January 30-February 2<sup>nd</sup>.

Keynes, John Maynard (1930). *A Treatise on Money*, Volume 1. New York: Hartcourt and Brace.

Kravis, I. B., Z. Kenessey, A. Heston and R. Summers (1975), *A System of International Comparisons of Gross Product and Purchasing Power* Baltimore: Johns Hopkins University Press.

Kravis, I. B., A. Heston and R. Summers (1978), "Aggregation Methods in International Comparisons of Real Product," *Eurostat News, Special Number*, Eurostat, Luxembourg.

\_\_\_\_\_ (1985), *World Product and Income*, Baltimore: Johns Hopkins University Press.

Pakes, Ariel, (2002), "A Reconsideration of Hedonic Price Indices with an Application to PCs," Brookings Workshop on Hedonic Price Indexes, Feb.

Prasada Rao, D.S. and M.P. Timmer (2000), *Multilateralisation of Manufacturing Sector Comparisons: Issues, Methods and Empirical Results*, SOM Research Report 00C61, University of Groningen.

Rao, D. S. Prasada (2001), "Weighted EKS and Generalized CPD Methods for Aggregation at Basic Heading Level and Above Basic Heading Level," Joint World Bank-OECD Seminar on Purchasing Power Applications, Washington D.C., [www.oecd.org/std/ppp/MTG2001/index.htm](http://www.oecd.org/std/ppp/MTG2001/index.htm), January 30-February 2<sup>nd</sup>.

Ryten, Jacob (1998), *The evaluation of the International Comparison Project (ICP): Summary of a report*, Cirencester and Ottawa, September.

Sergueev, Serguei (2001) "Measures of the similarity of the country's price structures and their practical application," Conference of European Statisticians, Consultation on the European Comparison Programme (Geneva, November)

United Nations and Eurostat, 1986, *World Comparisons of Purchasing Power and Real Product for 1980*, ST/ESA/STAT/SER F/42. New York.

United Nations Economic Commission for Latin America (1963), "A Measurement of Price Levels and the Purchasing Power of Currencies in Latin America, 1960-1962, Mar del Plata Argentina, May. E/CN.12/653.

Ward, M.(2002) "Price Ratios and the Estimation of Purchasing Power Parities" International Conference and Expert Group Meetings on the International Comparisons Program, World Bank, March 11-15.