Macroeconomic Dynamics, **13** (Supplement 2), 2009, 194–217. Printed in the United States of America.
doi:10.1017/S1365100509090129

RECENT DEVELOPMENTS IN THE INTERNATIONAL COMPARISON OF PRICES AND REAL OUTPUT

ROBERT J. HILL
University of Graz
and
University of New South Wales

T. PETER HILL

Three important current areas of research in the field of international comparisons are the construction of price indexes at the basic heading level, price indexes at higher levels of aggregation, and the linking of comparisons across regions. We consider recent innovations in each of these areas. These innovations have largely arisen out of International Comparisons Program (ICP) 2005, and hence we discuss them in this context. We give particular emphasis to the construction of price indexes at the basic heading level, because we believe it is here that the biggest problems lie. For example, the apparently anomalous results obtained for China and India in ICP 2005 can be traced back to problems at the basic heading level. We also highlight some inconsistencies in current ICP methodology and some promising areas for future research that warrant closer scrutiny in the next round of ICP.

Keywords: ICP, Basic Headings, Representativity Bias, Multilateral Methods, Linking Regions, Price Index

1. INTRODUCTION

The latest round of the International Comparisons Program (henceforth ICP 2005) is a huge undertaking coordinated by the World Bank in collaboration with the Organisation for Economic Co-operation and Development (OECD), Eurostat, the International Monetary Fund (IMF), and the United Nations. Its objective is to compare the purchasing power of currencies and real output across most countries in the world (146 countries participated in ICP 2005). ICP 2005 has led to a number of innovations in the methodology for making international comparisons of prices and real output.[1]

Perhaps the most pressing concern in the international comparisons literature is the problem of obtaining unbiased price indexes at the basic heading level (the lowest level of aggregation at which expenditure weights are available). The

T. Peter Hill is retired; formerly at OECD. Address correspondence to: Robert J. Hill, Department of Economics, University of Graz, Universitätsstrasse 15/F4, 8010 Graz, Austria; e-mail: robert.hill@uni-graz.at.

basic heading price indexes provide the building blocks from which the overall comparison is constructed. If these building blocks are biased or otherwise flawed, then everything that builds on them will be likewise tainted. ICP 2005 has been a source of some controversy, particularly with regard to the results obtained for China and India:

> The 2005 ICP, like earlier rounds, involved substantial revisions to previous data, most notably revising downwards the size of the Chinese (40 percent smaller) and Indian (36 percent) economies. [Deaton and Heston (2008)]

We believe that these seemingly anomalous results for China and India can be attributed to the fact that, as compared with most other countries, relatively more of the reported prices from which their basic heading price indexes were constructed were not representative. Deaton and Heston explain the problem as follows:

> Many of the qualities available in poorer countries are not available in higher income countries, while more of the qualities available of richer countries can also be found in poorer countries. . . . The consequence is that prices for the ICP were often collected in higher-end outlets, which has the effect of raising price levels of poorer countries. This was made more likely in 2005 than previously because of the much closer review of prices across countries so that, for example, international brands were priced in (say) China, because they were available, even if mainly in high-end outlets. To the extent this happened, it would have the effect of raising parities in poorer countries, making them appear to have less income and output than in fact they do. [Deaton and Heston (2008)]

This is clearly an issue that requires greater attention in the next round of ICP. It is possible to correct for representativity bias when constructing basic heading price indexes as long as the participating countries identify in a consistent manner the products on the list that are representative and those that are not. Unfortunately, this did not happen in ICP 2005 except in the Eurostat/OECD region. We discuss here in some detail why such adjustments are so important to the overall quality of the results.

We also consider the construction of price indexes at higher levels of aggregation. We revisit the debate over the relative merits of the GEKS [see Gini (1931), Eltetö and Köves (1964), and Szulc (1964)] and Geary (1958)–Khamis (1972) methods and of the IDB [see Iklé (1972)–Dikhanov (1994)–Balk (1996)] method, which has risen to prominence as a result of its use in ICP 2005 for comparisons in the African region. We then consider some extensions that combine the GEKS method with spanning-tree methods, which have the potential to either reduce the complexity or increase the reliability of the overall results.

Finally, we address regionalization. One of the defining aspects of ICP 2005 has been its reliance on regionalization. That is, the world is divided into six regions [Africa, Asia–Pacific, South America, Western Asia, Commonwealth of Independent States (CIS), and Eurostat/OECD]. Separate comparisons are made for each region. The comparisons in five of the regions are coordinated by the ICP Global office. The exception is Eurostat/OECD, which conducted its own 2005 comparison. These regions are then linked in a second stage using 18 so-called "ring" countries drawn from five of the regions. The CIS did not participate in the ring comparison. It is linked to Eurostat/OECD through Russia, which participated in both the Eurostat/OECD and CIS comparisons [see World Bank (2008)]. This process of regionalization has raised a number of conceptually challenging issues, not least of which is the problem of how best to link the regions. We consider the methods that have been proposed for calculating these between-region links.

In conclusion, we evaluate the progress that has been made as a result of ICP 2005 and suggest some improvements that can be made in future comparisons.

2. ESTIMATION OF PRICE INDEXES AT THE BASIC HEADING LEVEL

A fundamental distinction in international comparisons is between price indexes calculated at the basic heading level and at higher levels of aggregation.[2] A basic heading is a group of similar products defined within a general product classification.[3] Higher level aggregates consist of groups of basic headings. In general, expenditure data are not available at the level of individual products, the most detailed expenditure data available being for entire basic headings. Indeed some basic headings are de facto defined as the smallest groups of products for which expenditure data can be obtained. A price index for the products within a basic heading has therefore to be calculated from price data only. Indexes of this type are now described as elementary price indexes [see Diewert (2004a) and Hill (2004)].

A two-step procedure is employed in international comparisons in which the first step is to estimate elementary price indexes for each of the basic headings. The second step is to combine the basic heading indexes with the corresponding expenditure data to obtain higher level price indexes.

The academic literature has focused its attention predominantly on the aggregation methods used at the second stage. However, the estimation of the elementary price indexes is at least as important as the way in which they are subsequently aggregated. Their estimation raises important methodological issues that have been neglected in most of the academic literature. A number of important methodological innovations have been introduced in the joint Eurostat/OECD Purchasing Power Parity (PPP) Program that are relevant for other regions also. These concern not only the formulas used to estimate the basic heading price indexes but also the criteria and methods used to select the products whose prices are collected.

2.1. Price Collection for International Comparisons

There are 155 basic headings in ICP 2005. Some may contain almost countless numbers of products depending on how finely the products, are defined and classified. For purposes of both temporal indexes and international comparisons, it is feasible to collect prices for only a small number of products in each basic heading. The products are selected purposively, and the choice of products can have a significant influence on the resulting elementary price indexes. The ICP compares national average prices so that for each product prices have to be collected in a sample of different outlets throughout the country.[4] Product lists that statistical offices draw up independently of each other for their own temporal price indexes inevitably differ from country to country. If there is very little overlap between them, they are not suitable for international comparisons.

For international comparisons, prices must be collected for the same products in different countries. A common list of products must therefore be worked out in advance by the project managers and the countries. To ensure that price collectors in different countries working independently of each other collect prices for exactly the same products, the products must also be specified in considerable detail: i.e., be tightly defined. In ICP 2005 a completely new set of Standard Product Descriptions, or SPDs, were developed, which provide detailed checklists of the characteristics of each of the products on the common list.

International comparisons may be bilateral or multilateral. Even if the objective is to calculate a set of multilateral comparisons, bilateral indexes may be calculated first and transitivity subsequently imposed on them. This is the approach used both for basic headings and higher level aggregates by Eurostat and the OECD. It is also used at an aggregate level in ICP 2005. It is appropriate to consider it first.

2.1.1. Bilateral elementary price indexes. International price indexes are often called purchasing power parities, or PPPs. At the level of an individual product a PPP is simply the ratio of the national average prices of the product in two countries denominated in their own currencies. It is the rate at which a given amount of money must be converted from one currency to the other in order to ensure that it purchases the same quantity of the product in the two countries: hence the term *purchasing power parity*. At the level of an individual product, the price ratios, or PPPs, between different pairs of countries are transitive.

Patterns of relative prices, and quantities, can vary considerably from country to country and especially between countries at different levels of economic development or with different cultures and climates. The price ratios for individual products also vary therefore.

In a bilateral comparison the elementary price index for a basic heading is defined as the geometric mean of the price ratios of the products within the basic heading. This type of elementary index is now described in the price index literature as a Jevons index:

$$p_{jk}^{n} = \prod_{m=1}^{M} \left(\frac{p_{km}}{p_{jm}} \right)^{1/M},$$

where j and k denote a pair of countries, n denotes a particular basic heading, and $m = 1, \ldots, M$ indexes the set of products over which the basic heading price index is calculated.

Jevons indexes based on purposive samples of products are widely used in temporal indexes such as consumer price indexes (CPIs) and producer price indexes (PPIs). They are preferred over other possible types of elementary index on both axiomatic and economic grounds.[5]

Jevons indexes are transitive when all countries supply prices on exactly the same list of products.

In practice, however, it would not be desirable to confine the list of products to those that can be priced in every country. Such a list would be too restrictive and might not be representative of most of the countries in the comparison. In consequence, there are usually gaps in the price table and hence this transitivity property is rarely satisfied.

2.1.2. Relative prices and balanced lists of products. In a bilateral comparison, products m that are relatively cheap in country B compared with country A can be identified simply by ranking the price ratios p_{Bm}/p_{Am} in ascending order. Suppose that such a list is constructed over the whole universe of products in a basic heading. The products in the first half of the list are relatively cheap in country B while the remainder are relatively expensive.[6] A small purposively selected sample or basket of products may well be unbalanced in the sense that the observations may be drawn more from one or other side of the distribution. If most of the products selected in the basket are relatively cheap in B, the Jevons index for the basket is less than the Jevons index for the basic heading as a whole. That is, the reported Jevons index has a downward bias. The key question becomes how the sample of products is determined in practice.

In a simple bilateral comparison, the two countries should collaborate in drawing up the sample of products for price collection. Each should try to ensure that the common list of products is balanced. As the relative prices are not actually known, price experts in the two countries have to use their knowledge of markets to make a judgment about which products are likely to be relatively cheap or expensive. Balance may be difficult to achieve, but at least gross imbalances should be avoidable.

Elementary index bias resulting from the use of unbalanced lists will not be eliminated at higher levels of aggregation by the use of superlative indexes. Higher level Laspeyres and Paasche indexes are weighted arithmetic and weighted harmonic averages of the same elementary price indexes using basic heading expenditures as weights. If the elementary price indexes tend to be systematically biased in the same direction, the Laspeyres and the Paasche will both have similar biases and hence also the Fisher index.

2.1.3. Product lists for multilateral comparisons. As already noted, multilateral price indexes may be estimated either by calculating a complete set of bilateral comparisons and imposing transitivity on them, if necessary, or by calculating a transitive set of multilateral price indexes directly. In either case, a common product list has to be constructed for use by all the countries in the group. The total number of bilaterals among a group of K countries is $K(K-1)/2$. It would be quite unrealistic to expect that number of separate product lists to be drawn up, each involving a significant amount of work for the pair of countries concerned. A proliferation of product lists would require each country to collect a much larger number of prices, which again is unrealistic. In practice, therefore, each bilateral comparison made within the framework of a set of multilateral comparisons uses the same common list. The list used will be different from one that would be used in an independent bilateral comparison. The comparison will therefore not be fully characteristic of the two countries concerned as the list will have been influenced to some extent by third countries.

If the proportions of relatively cheap and expensive products on the common multilateral product list are not the same in a country, the list is unbalanced from the viewpoint of that country. However, such imbalance does not necessarily introduce bias. Provided that the ratio of relatively cheap to relatively expensive products on the list is the same in the two countries compared, the simple Jevons index between them will not be biased. In this case, the list is said to be equally unbalanced for the two countries. In the special case in which the ratio equals unity in both countries, the list is balanced for both countries and the Jevons is unbiased.

It is not possible for the list to be equally unbalanced for every pair of countries in a group. If the proportion of relatively cheap products is greater than half in some countries, there must be some other countries in which it is less than half. It may happen that for an entire subgroup of countries most of the products on the list are relatively cheap while for another subgroup they are relatively expensive. In this case, the simple Jevons indexes between countries in the same group may be unbiased but those between countries belonging to different subgroups will be biased. In a situation of this kind it may be better to adopt a regional approach and split the initial group or region into two separate regions and link the two sets of regional results. Only in the unlikely event that the list is balanced for every country would the entire set of Jevons indexes be unbiased.

A further complication is that countries typically do not provide prices for all the products on the common list. In practice, the Jevons index has to be calculated for the list of products for which prices are reported by both countries. As the missing prices tend to differ from country to country, this list of products priced is less likely to be equally unbalanced for both countries than the original common list. The risk of bias in the simple Jevons indexes is likely to be increased by missing prices.

The use of a common product list in a multilateral comparison will almost inevitably result in the list containing more of the representative products of some countries than others. This may create a bias analogous to the Gerschenkron effect (see Section 3.1). That is, prices will tend to be overestimated and incomes underestimated in the countries for which the list is least representative.[7]

2.1.4. Representativity. Eurostat has always used the bilateral approach in its PPP program. Over the years it has developed procedures for combating possible biases of the kind just described. These require countries to distinguish between

https://doi.org/10.1017/S1365100509090129 Published online by Cambridge University Press

representative and unrepresentative products. The exact meaning of "representative" in this context is not entirely clear, but the purpose of the distinction is. The Eurostat-OECD Methodological Manual on PPPs [Roberts (2006)] states that

> Representative products normally have a lower price level than unrepresentative products and, if this is not taken into account when calculating the PPPs for a basic heading, the PPPs will be biased. To avoid this, participating countries are required to identify which of the products they have priced within a basic heading are representative when reporting their prices. [Roberts (2006, p. 5)]

If products are relatively cheap they are likely to be purchased in relatively large quantities. Representative products should therefore generally be relatively popular. Unrepresentative products tend to be purchased in relatively small quantities and have relatively high prices.

One reason for distinguishing representative from unrepresentative products is to improve the product lists. The preparation of suitable lists is a lengthy process that requires the active participation of the countries as well as the program managers. This may require several meetings between country experts and the program managers. At the minimum it is necessary to ensure that each list includes enough of the representative products of every country.

However, the main reason for distinguishing representative from unrepresentative products is that this information can be used to obtain unbiased estimates of the elementary price indexes when the lists of products for which prices are reported are unbalanced for most countries. Countries participating in the Eurostat–OECD program

> are required to indicate which of the products they have priced are representative when reporting their prices. Representative products are designated by a "representativity indicator" . . . an asterisk (∗) and . . . are called "asterisk products." [Roberts (2006, paragraph 102)]

Ways in which the information can be utilized are explained in the following section. Although the distinction between representative and unrepresentative products was pioneered by Eurostat using a bilateral approach, it is equally important when a multilateral method such as the country-product-dummy (CPD) is used.

As there is usually little or no hard information about prices, quantities, and expenditures within a basic heading, distinguishing representative from unrepresentative products is partly a matter of judgment for country price experts. Some countries have difficulty in making the distinction, especially when asked to do so for the first time. The Eurostat-OECD group of countries have become familiar with the distinction over a number of years, but for countries in other regions the distinction was new in ICP 2005.

2.2. The Eurostat Method

2.2.1. Jevons indexes.* Starting in the 1980s, Eurostat refined its methodology for calculating the bilateral indexes [see Eurostat (1983)]. This refinement was made possible by the fact that Eurostat started asking countries to identify each product in the list for each basic heading as either representative or not representative for that country. This extra information allows two separate Jevons indexes to be calculated for each bilateral comparison, one based on the list of products identified as representative for country j and the other based on the list of products identified as representative for country k. The Jevon* index is obtained by taking the geometric mean of these two indexes:

$$p_{jk}^{R_{rj-k},n} = \prod_{m_{rj-k}=1}^{M_{rj-k}} \left(\frac{p_{km_{rj-k}}}{p_{jm_{rj-k}}} \right)^{1/M_{rj-k}},$$

$$p_{jk}^{R_{rk-j},n} = \prod_{m_{rk-j}=1}^{M_{rk-j}} \left(\frac{p_{km_{rk-j}}}{p_{jm_{rk-j}}} \right)^{1/M_{rk-j}},$$

$$\text{Jevons}^*: \; p_{jk}^{*,n} = \sqrt{p_{jk}^{R_{rj-k},n} \times p_{jk}^{R_{rk-j},n}}, \tag{1}$$

where $m_{rj-k} = 1, \ldots, M_{rj-k}$ indexes the set of products that are representative in country j and that are also priced in country k (although are not necessarily marked as representative in the latter).[8] $p_{jk}^{R_{rj-k},n}$ is a price index for basic heading n calculated over the products that are representative in country j and also priced in country k.

The Jevons* indexes are essentially examples of Törnqvist indexes, with shares $s_{jm} = 1/M_{rj-k}$ for representative products in country j and $s_{jm} = 0$, for non-representative products, and likewise $s_{km} = 1/M_{rk-j}$ for representative products in country k and $s_{km} = 0$ for nonrepresentative products. The Törnqvist index is defined:

$$p_{jk}^{n} = \prod_{m=1}^{M} \left[\left(\frac{p_{km}}{p_{jm}} \right)^{(s_{jm}+s_{km})/2} \right],$$

where $m = 1, \ldots, M$ indexes the list of products included in the basic heading. These Törnqvist indexes are then transitivized using the GEKS formula:

$$\frac{p_{kn}}{p_{jn}} = \prod_{l=1}^{M} \left(\frac{p_{lk}^{n}}{p_{lj}^{n}} \right)^{1/M}. \tag{2}$$

2.2.2. Jevons-S indexes. The Jevons-S index is a slight improvement on Jevons* proposed by Sergeev (2003). We can rearrange (1) as follows:

$$p_{jk}^{R_{rj-k},n} = \left[\prod_{m_{rj-uk}=1}^{M_{rj-uk}} \left(\frac{p_{km_{rj-uk}}}{p_{jm_{rj-uk}}} \right)^{1/M_{rj-k}} \right] \left[\prod_{m_{rj-rk}=1}^{M_{rj-rk}} \left(\frac{p_{km_{rj-rk}}}{p_{jm_{rj-rk}}} \right)^{1/M_{rj-k}} \right],$$

https://doi.org/10.1017/S1365100509090129 Published online by Cambridge University Press

$$p_{jk}^{R_{rk-j},n} = \left[\prod_{m_{rk-uj}=1}^{M_{rk-uj}} \left(\frac{p_{km_{rk-uj}}}{p_{jm_{rk-uj}}}\right)^{1/M_{rk-j}}\right]\left[\prod_{m_{rj-rk}=1}^{M_{rj-rk}} \left(\frac{p_{km_{rj-rk}}}{p_{jm_{rj-rk}}}\right)^{1/M_{rk-j}}\right],$$

$$\text{Jevons}^*\text{:}\quad p_{jk}^{*,n} = \sqrt{p_{jk}^{R_{rj-k},n} \times p_{jk}^{R_{rk-j},n}} = \left[\prod_{m_{rj-uk}=1}^{M_{rj-uk}} \left(\frac{p_{km_{rj-uk}}}{p_{jm_{rj-uk}}}\right)^{1/(2M_{rj-k})}\right]$$

$$\times \left[\prod_{m_{rj-rk}=1}^{M_{rj-rk}} \left(\frac{p_{km_{rj-rk}}}{p_{jm_{rj-rk}}}\right)^{1/(2M_{rj-k})+1/(2M_{rk-j})}\right]\left[\prod_{m_{rk-uj}=1}^{M_{rk-uj}} \left(\frac{p_{km_{rk-uj}}}{p_{jm_{rk-uj}}}\right)^{1/(2M_{rk-j})}\right], \quad \textbf{(3)}$$

where $m_{rj-uk} = 1, \ldots, M_{rj-uk}$ indexes the set of products that are representative in country j and unrepresentative in country k; $m_{rk-uj} = 1, \ldots, M_{rk-uj}$ indexes the set of products that are representative in country k and unrepresentative in country j; and $m_{rj-rk} = 1, \ldots, M_{rj-rk}$ indexes the set of products that are representative in both countries j and k.

It can be seen that the basic heading price index is a weighted average of three groups of products. These groups are as follows:

Group 1: Products that are representative for country j but unrepresentative for k.
Group 2: Products that are representative for both countries j and k.
Group 3: Products that are representative for country k but unrepresentative for j.

The weights for each group obtained by multiplying the number of observations in each group by the weight given to each individual observation in each group are as follows:

$$\text{Group 1: } \frac{M_{rj-uk}}{2M_{rj-k}}, \qquad \text{Group 2: } \frac{M_{rj-rk}(M_{rj-k} + M_{rk-j})}{2(M_{rj-k} \times M_{rk-j})},$$

$$\text{Group 3: } \frac{M_{rk-uj}}{2M_{rk-j}}. \quad \textbf{(4)}$$

Using the fact that $M_{rj-rk} = M_{rk-rj}$ and that $M_{rj-uk} + M_{rj-rk} = M_{rj-k}$, it can be shown that these weights sum to one.

In general, there is no reason to expect that the Group 1 and 3 weights are equal. This is potentially problematic because, other things equal, representative products tend to be cheaper than unrepresentative products. In other words, the price index $p_{jk}^{*,n}$ will tend to have an upward bias if calculated using only products from Group 1 and a downward bias if calculated using only products in Group 3. To prevent bias in the index, it is important, therefore, that these two groups receive equal weight in the price index formula. This is achieved by recalibrating the weights for Groups 1 and 3 so that they both equal

$$\frac{M_{rj-uk}}{4M_{rj-k}} + \frac{M_{rk-uj}}{4M_{rk-j}}.$$

This adjustment leaves the weight on Group 2 unaffected.

The Jevons-S method departs slightly from the above in that it counts each observation in Group 2 twice, on the grounds that these estimates should be unbiased and hence more reliable [again see Sergeev (2003)]. This increases the weight on Group 2 to

$$\frac{2M_{rj-rk}(M_{rj-k} + M_{rk-j})}{2(M_{rj-k} \times M_{rk-j}) + M_{rj-rk}(M_{rj-k} + M_{rk-j})}.$$

The weights on Groups 1 and 3 are correspondingly scaled down by a proportion:

$$1 + \frac{M_{rj-rk}(M_{rj-k} + M_{rk-j})}{2(M_{rj-k} \times M_{rk-j})}.$$

Again the price indexes must be transitivized using the GEKS formula in (2).

2.3. The CPD Method

The CPD method, first proposed by Summers (1973), calculates the price index for a basic heading n for all countries simultaneously, and hence there is no need to apply the GEKS transitivization formula. The CPD model estimates the following regression equation:

$$\ln p_{km} = \kappa + \sum_{\mu=2}^{M} \alpha_\mu x_\mu + \sum_{j=2}^{K} \beta_j y_j + \varepsilon_{km}, \tag{5}$$

where p_{km} denotes the price of product m in country k, x_μ denotes a product dummy variable that equals 1 if $m = \mu$ and zero otherwise, y_j denotes a country dummy variable that equals 1 if $k = j$ and zero otherwise, and ε_{km} denotes a random error term. The κ, α_m, and β_k parameters can be estimated by least squares. Exponentiating the estimated β_k parameters, we obtain the price index for a basic heading for each country.[9–11]

The estimated error term $\hat{\varepsilon}_{km}$ in equation (5) is equal to the difference between the log of the actual price p_{km} and the log of the imputed price obtained from the CPD model. This is the log of the ratio of the actual to the imputed price. This ratio measures the relative price of product m in country k. If $\hat{\varepsilon}_{km}$ is negative, this implies that product m is relatively cheap in country k compared with other countries in the group.[12]

By construction it will be the case that for each country k that $\sum_{m=2}^{M} \hat{\varepsilon}_{km} = 0$. It is useful at this point to distinguish between two product lists. The first is the ideal list consisting of the whole universe of products in a particular basic heading. The second is the actual and much smaller list of products used in practice. Suppose now that it were possible to run a CPD regression on the ideal product list. By definition, the resulting price indexes $\exp(\hat{\beta}_k^*)$ should be unbiased. (Here the asterisk superscript denotes a price index calculated using the ideal product list.) How will the price indexes calculated using the actual product list $\exp(\hat{\beta}_k)$

https://doi.org/10.1017/S1365100509090129 Published online by Cambridge University Press

compare? This depends on whether the products deleted when moving from the ideal to actual lists predominantly had negative or positive estimated error terms in the original ideal CPD regression. Here we will index the deleted products by $m = m^d, \ldots, M$. If $\sum_{m=m_d}^{M} \hat{\varepsilon}_{km} < 0$ in the ideal CDP regression, then it follows that $\exp(\hat{\beta}_k) > \exp(\hat{\beta}_k^*)$. In other words, the actual price index for country k will have an upward bias (at least relative to the base country). The bias goes in the opposite direction when $\sum_{m=m_d}^{M} \hat{\varepsilon}_{km} > 0$. The actual list is perfectly balanced, and hence the CPD price indexes are free of bias only if $\sum_{m=m_d}^{M} \hat{\varepsilon}_{km} = 0$ for all k. This condition is unlikely to be even approximately satisfied.

Just as it is possible to correct for biases in simple Jevons indexes by using information about representativity, it is possible to correct for biases in the CPD estimates by extending the method to include representativity. The extended CPD method [first proposed by Cuthbert and Cuthbert (1988)] is now referred to as the country-product-representative-dummy (CPRD) method [see Hill (2007a)]. It simply adds a dummy variable to the model as follows:

$$\ln p_{km} = \kappa + \sum_{\mu=2}^{M} \alpha_\mu x_\mu + \sum_{j=2}^{K} \beta_j y_j + \gamma_i z_i + \varepsilon_{km},$$

where z_i is a dummy that equals 1 if product m is representative in country k and zero otherwise.

Assuming that a representative product is relatively cheap, the error term, $\hat{\varepsilon}_{km}$, for a product that is representative in country k in the simple CPD for the basic heading as a whole will be negative. If representative products can be identified, this information can be utilized to correct for imbalances between the proportions of representative and unrepresentative products on the list for different countries. In effect, either the prices of representative products can be adjusted upward by a representativity factor or the prices of unrepresentative products can be adjusted downward. The CPRD method estimates the adjustment factor simultaneously with the product and country factors.

The use of a selected list of products, whether selected randomly or purposively, will not introduce bias into the estimated basic headings PPPs in the simple CPD model when the errors are random. In practice, however, there tend to be systematic differences in relative prices between countries, especially between countries at different levels of economic development, so that the errors in the simple CPD model are almost certainly not random. For example, for certain types of products, richer countries are more likely to have positive errors and poorer countries negative errors. For other products, the reverse will be true.[13] Under this more realistic scenario, the use of selected lists is likely to cause bias. The CPRD method explicitly allows for the fact that the relative price of a product may be high in some countries and low in others and goes a long way toward eliminating potential biases resulting from specification errors in the simple CPD method.

When the price table for a basic heading is complete (i.e., all products in the list are priced by all countries), it turns out that the Eurostat–Jevons basic heading price indexes are identical to their CPD counterparts [see Rao (2004)].[14,15] The equivalence of Jevons and CPD is of limited practical relevance, because there are typically large gaps in the price table. The Jevons-S imbalance between the Group 1 and 3 weights in (4) is a direct consequence of an incomplete price table. Hence there is no difference between Jevons* and Jevons-S when the price table is complete. In this sense, Jevons-S can be viewed as a correction for missing data. It does not necessarily follow, however, from this that Jevons-S is superior to CPRD. The CPD and CPRD methods naturally correct for missing data.[16]

At its meeting in September 2004, the ICP 2005 Technical Advisory Group

> recommended that regions should use the CPRD method to estimate basic heading PPPs. Of course, the method can only be implemented satisfactorily if the countries within a region are able to identify representative products correctly. [Hill (2007a)]

We endorse this position. Unfortunately,

> Economies in the Asia-Pacific, Africa, Western Asia, and South America regions that either had not participated in an international comparison for an extended period or had never participated had difficulty applying the representativity concept, therefore, it was not used in their intraregional comparisons. [World Bank (2008, p. 185)]

Actually, this statement is not quite correct. It seems that South America did use CPRD, even though the signs on the representativity coefficients were more or less random, suggesting that representativity was not identified in a consistent manner across countries [see Diewert (2008)]. The failure of the Asia–Pacific, Africa, Western Asia, and South America regions to effectively implement the CPRD method means that some of the estimated basic heading PPPs in these regions are exposed to a significant risk of bias. These biases can explain the apparently anomalous results obtained for China and India referred to above.

The other two regions (CIS and Eurostat/OECD) used the Jevons-S method, which as already noted also requires the identification of representative products. The Eurostat/OECD countries, however, have been identifying representative products for many years and hence presumably long ago resolved any issues of inconsistency. By implication, the Eurostat/OECD results are likely to be more reliable than those obtained for the other regions.

We are nevertheless inclined in principle to favor the CPRD method over the Jevons-S method due to its greater simplicity, transparency, more straightforward treatment of missing prices, and the fact that it provides standard errors on the basic heading price indexes.[17] Whichever method is used, for the sake of internal consistency it would be preferable that all regions use the same method to compute the basic heading price indexes.

3. MULTILATERAL AGGREGATION ABOVE BASIC HEADING LEVEL

3.1. Average Price Methods

A large number of multilateral methods have been proposed in the literature for computing price indexes above basic heading level [see Hill (1997)]. The two methods that have attracted the most attention are Geary (1958)–Khamis (1972) and GEKS. Geary–Khamis is an example of an average price method. It calculates the multilateral price index for country k relative to country j as follows:

$$P_{jk} = \frac{\sum_{n=1}^{N} p_{kn}q_{kn}}{\sum_{n=1}^{N} p_{Xn}q_{kn}} \frac{\sum_{n=1}^{N} p_{Xn}q_{jn}}{\sum_{n=1}^{N} p_{jn}q_{jn}} = \frac{P_{Xk}^{P}}{P_{Xj}^{P}},$$

where X is an artificial average country with prices for each basic heading $n = 1, \ldots, N$ given by

$$p_{Xn} = \sum_{k=1}^{K} \left(\frac{q_{kn}}{\sum_{j=1}^{K} q_{ji}} \frac{p_{kn}}{P_{Xk}^{P}} \right),$$

and P_{Xk}^{P} is a Paasche price index defined as follows:

$$P_{Xk}^{P} = \frac{\sum_{n=1}^{N} p_{kn}q_{kn}}{\sum_{n=1}^{N} p_{Xn}q_{kn}}.$$

The Geary–Khamis price indexes and average price vector are obtained by solving this system of $N + K - 1$ simultaneous equations (i.e., $K - 1$ multilateral price indexes and N average basic heading prices). A sufficient condition for the system to have a unique strictly positive solution is that all prices and quantities are strictly positive [see Khamis (1972)].

Average price methods can be thought of as making star comparisons between each country and an artificial average country X, with the artificial country at the center of the star and each comparison made using the Paasche price index formula with country X as the base. They differ in the formula for determining the average prices p_{Xn}.

One attractive feature of average price methods is that they are additive. That is, their quantity indexes add up over different levels of aggregation when measured in value terms. Additivity is very useful in comparisons made over varying levels of aggregation as, for example, in national accounts comparisons, and indeed in ICP itself.

The main drawback of additive methods is that they are subject to substitution bias (otherwise known as the Gerschenkron effect). In most data sets the consumer substitution effect dominates the producer substitution effect and hence Laspeyres price indexes systematically exceed their Paasche counterparts (in the absence of strong offsetting income effects). The Paasche indexes P_{Xk}^{P} that underpin average price methods therefore have a downward bias. The magnitude of this bias is an increasing function of the dissimilarity between the average price vector p_X

https://doi.org/10.1017/S1365100509090129 Published online by Cambridge University Press

and the price vectors of the countries p_k [see Hill (2000)]. The implication is that countries with relative prices that differ substantially from the average prices will appear richer than they really are. The Geary–Khamis average price vector usually approximates most closely the price vectors of the richer countries in the comparison. It follows that Geary–Khamis tends to systematically underestimate differences in per capita income levels across countries [see Dowrick and Quiggin (1997) and Hill (2000)].

An additive alternative to Geary–Khamis is the IDB method. The IDB method calculates the average basic heading prices as follows:

$$p_{Xn} = \sum_{k=1}^{K} \left(\frac{q_{kn}/Q_{Xk}^{L}}{\sum_{j=1}^{K} q_{ji}/Q_{Xj}^{L}} \frac{p_{kn}}{P_{Xk}^{P}} \right),$$

where Q_{Xk}^{L} is a Laspeyres quantity index defined as follows:

$$Q_{Xk}^{L} = \frac{\sum_{n=1}^{N} p_{Xn} q_{kn}}{\sum_{n=1}^{N} p_{Xn} q_{Xn}}.$$

It is not necessary to calculate the average q_X because it drops out of the solution. Again, a sufficient condition for the system to have a unique strictly positive solution is that all prices and quantities are strictly positive [see Balk (1996)]. Weaker conditions are considered in Diewert (2008).

The importance of the IDB method has recently been boosted by the fact that it was used by Africa in ICP 2005. The key difference between Geary–Khamis and IDB is that the latter gives equal weight to all countries in the average price formula. It follows that it should not have any systematic tendency to either overestimate or underestimate differences in per capita income across countries. It is for this reason that it was preferred to Geary–Khamis in ICP 2005. This, however, does not imply that IDB is unaffected by substitution bias. The distortions created by substitution bias are now simply harder to predict.

3.2. GEKS and Spanning-Tree Methods

Five of the six regions in ICP 2005 (the exception being Africa) use the GEKS method. The building blocks of the GEKS method are bilateral comparisons between all possible pairings of countries. The GEKS method essentially generates K sets of results, putting each country in turn at the center of a star and then computing Fisher indexes (or some other index) between it and each of the other countries in the comparison. A geometric average is then taken of these K sets of results.

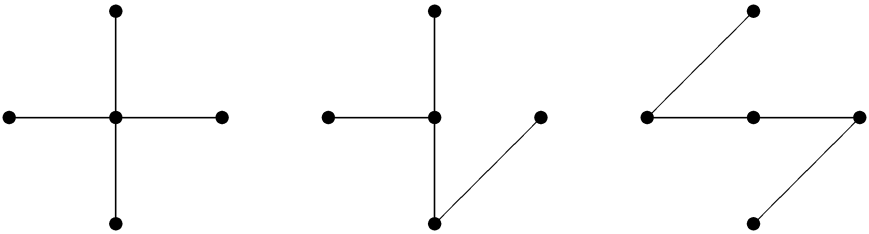


FIGURE 1. Examples of spanning trees for $K = 5$.

In ICP 2005 these bilateral comparisons are made using the Fisher price index formula (the geometric mean of Paasche and Laspeyres) defined as follows:

$$P_{jk}^F = \sqrt{P_{jk}^P \times P_{jk}^L} = \sqrt{\frac{\sum_{n=1}^N p_{kn}q_{kn}}{\sum_{n=1}^N p_{jn}q_{kn}} \frac{\sum_{n=1}^N p_{kn}q_{jn}}{\sum_{n=1}^N p_{jn}q_{jn}}}.$$

These Fisher indexes are then transitivized as follows:

$$P_{jk} = \prod_{l=1}^{K} \left[\left(\frac{P_{lk}^F}{P_{lj}^F} \right)^{1/K} \right].$$

The GEKS method is not additive. However, it also is not affected by substitution bias.

Using all possible pairings of bilateral comparisons has two associated disadvantages. First, some bilateral comparisons are more reliable than others. An indication of reliability can be obtained from the spread between Paasche and Laspeyres price indexes. The bigger the spread, the more sensitive the result is to the choice of bilateral price index formula, and hence the less reliable is the result.[18] Hence it may be possible to improve the reliability of the overall comparison by focusing on only the more reliable bilateral comparisons. The second disadvantage of using all possible bilateral comparisons is that all countries have to use the same product lists for each basic heading. If instead the overall comparison was broken up into smaller blocks, then each block could have its own product list for each basic heading and even its own list of basic headings, both of which could be tailored more closely to the expenditure patterns of the countries in that block, thus improving characteristicity.

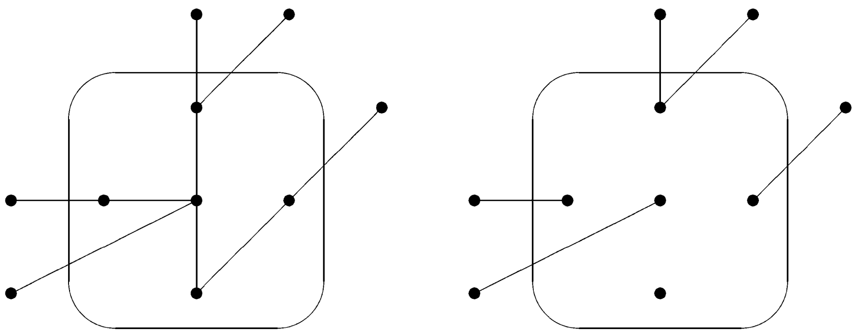
The minimum-spanning-tree (MST) method of Hill (1999) offers a useful starting point for variants on the GEKS method. In the graph theory literature, a spanning tree is defined as a graph that connects a set of vertices without creating any cycles. That is, there is one and only one path between each pair of vertices. Examples of spanning trees are shown in Figure 1. In the context of international comparisons, each country is represented by a vertex and bilateral comparisons (e.g., using the Fisher index) by edges linking vertices.[19] By using a spanning tree, all internal inconsistencies arising from the intransitivity of the bilateral price index formula are removed from a comparison. The spanning tree describes exactly how the multilateral indexes should be constructed from the matrix of bilateral indexes.

In a comparison between K countries there are K^{K-2} possible spanning trees that could be used, each of which will generate a different set of multilateral price indexes even if they all use the same price index formula to make the bilateral comparisons. Hill (1999) argued for using the MST, with the weights on the bilateral comparisons (edges) given by the Paasche–Laspeyres spreads [or one of the measures proposed by Diewert (2002)]. Kruskal's algorithm can be used to compute the MST. It starts by selecting the two bilateral comparisons (edges) with the smallest weight and thereafter selects the next smallest edge subject to the constraint that its inclusion does not create a cycle in the graph. If it does create a cycle, this edge is skipped. This process continues until $K - 1$ edges have been selected. At this point it is no longer possible to add any more edges without creating a cycle. The resulting spanning tree is the MST. It is the spanning tree with the minimum sum of weights [see, for example, Wilson (1985)].

One advantage of using a spanning tree (minimum or otherwise) is that it breaks the multilateral comparison into $K - 1$ bilateral comparisons. In principle, each of these bilateral comparisons could have its own product list, thus increasing the characteristicity of the comparison and reducing its logistical complexity. Also, if these $K - 1$ bilaterals are chosen for their reliability, this should further enhance the quality of the multilateral comparison.

One problem with the MST method is that it does not take account of how well resourced the individual national statistical offices are. This point has been made, for example, by Dikhanov in e-mail correspondence (September 25, 2008). Diewert, in an e-mail response on the same day, suggested a solution to this problem. Before constructing a MST, the countries could be sorted into two groups, consisting of those with well resourced statistical offices and those with less well resourced offices. Henceforth, we will refer to the former group as the core.[20] Diewert suggests computing a MST using only the core countries. Each noncore country is then linked to the core by its minimum weight edge. For example, suppose that there are 10 core countries and that country Z is a noncore country. Then there are 10 possible ways that country Z can be linked to the core. All that is required is to find which is the best of these 10 possible links (e.g., the one with the smallest Paasche–Laspeyres spread). This approach ensures that noncore countries do not play a pivotal role in the spanning tree. The noncore countries will all lie on the periphery. This method is illustrated in Figure 2(a). The core countries lie within the oval, and the noncore countries are outside the oval.

A second disadvantage of the MST method is that it tends to lack robustness from one cross section to the next [see Hill (1999)]. The robustness problem can be countered in two ways. First, it can be argued that what really matters is not so much the robustness of the MST itself, as the sensitivity of the resulting multilateral



(a) Core comparison using MST (b) Core comparison using GEKS

FIGURE 2. Core and noncore countries.

price indexes to changes in the spanning tree. For a reasonably homogeneous group of countries Fisher and Törnqvist indexes will be nearly transitive [see Deaton and Heston (2008)], and hence the multilateral price indexes will not be that sensitive to the choice of spanning tree.[21]

The second way of dealing with the robustness problem is to use a hybrid method that combines elements of GEKS with elements of the MST method. For example, the core comparison could be made using GEKS rather than a spanning tree. That is, bilateral comparisons between all pairings of countries in the core are used, which are then transitivized using the GEKS formula. Each noncore country is then linked with a core country in the manner described above and outlined in Figure 2(b). The problem with this method is that bilateral comparisons between some of the core countries could still have large Paasche–Laspeyres spreads, thus reducing the reliability of the overall results.

Another possible hybrid method combines the GEKS method with the concept of a shortest path. Before applying the GEKS transitivization formula, the method begins by calculating the shortest path between each pair of countries, calculated, say, by the chained Paasche–Laspeyres spreads along each path. For example, if there are four countries, A, B, C, and D, there are five possible paths (in which each edge features not more than once) from A to D. These are $A-D$, $A-B-D$, $A-C-D$, $A-B-C-D$, and $A-C-B-D$. We would select whichever of these five paths has the smallest associated chained Paasche–Laspeyres spread. In some cases the shortest path (which can be easily calculated using the shortest path algorithm) will not be the direct comparison (i.e., $A-D$). For example, it may be the case that the shortest path between Sweden and Greece is via Germany. If the Fisher index is used to make the bilateral comparisons, it follows that we would replace the direct Fisher between Sweden and Greece with a chained Fisher index (via Germany). The GEKS transitivization formula is then applied to the adjusted matrix of bilateral Fisher indexes (i.e., where some of the direct

https://doi.org/10.1017/S1365100509090129 Published online by Cambridge University Press

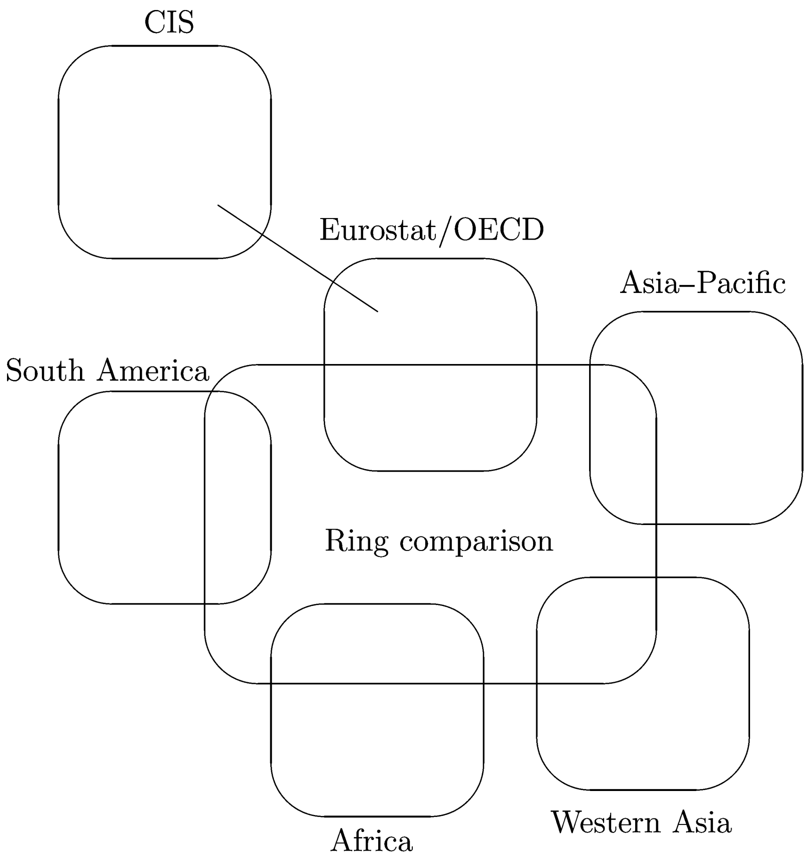


FIGURE 3. The two-stage structure of ICP 2005.

Fishers have been replaced by chained Fishers). The advantage of this method is that the most egregious bilaterals are replaced by chained bilaterals with smaller Paasche–Laspeyres spreads prior to the application of the GEKS formula.

The application of spanning-tree methods, particularly in combination with GEKS, is a promising area for future research and deserves serious consideration in the next round of ICP.

4. REGIONALIZATION OF ICP 2005

ICP 2005 is a huge undertaking involving 146 countries. As a result of the sheer scale and complexity of the project, ICP 2005 has been broken up into two stages. The two stages are illustrated in Figure 3. In stage 1, the countries are divided into six regional blocks. The regions are Africa, Asia–Pacific, South America, Western Asia, CIS, and Eurostat/OECD. Separate comparisons are then made for each block. In stage 2, these regional results are linked to generate

https://doi.org/10.1017/S1365100509090129 Published online by Cambridge University Press

the overall global results. This is achieved by means of a comparison between 18 so-called ring countries drawn from five of the regions, with at least two ring countries in each region [see Hill (2007b)]. The CIS is linked to Eurostat/OECD through Russia, which participated in both the Eurostat/OECD and CIS comparisons.

Manageability, however, is not the only reason for this two-stage approach. A second factor was the desire to maximize characteristicity. By breaking the comparison up into separate regional blocks, each region could price its own list of products that are representative of its region.[22]

Ideally, the two stages should be harmonized. This raises the question of how best to construct price indexes in this context. A similar problem was encountered previously by the OECD, as a result of Eurostat's requirement of fixity in the results for the European Union. This essentially requires the OECD to split itself into two regions (i.e., the EU and the OECD excluding the EU) in its comparisons. The OECD's solution to this problem is not very satisfactory, in that it does not treat the two regions symmetrically [see Hill and Hill (2007) for further details]. Symmetric treatment of regions, by contrast, is an essential requirement of ICP 2005.

This two-stage problem has provided a stimulus for new research. The problem arises because there are at least two ring countries in each region. This was a deliberate strategy in ICP 2005 so as to increase the robustness of the between-region price indexes. By contrast, if each region had only one ring country, the construction of regional price indexes would be straightforward. The same method could be applied to link the regions as is used in stage 1 to make the within-region comparisons.

Solutions to the problem of computing between-region price indexes when each region has two or more ring countries have been proposed by Diewert (2004b, 2008) and Hill (2005) [see also Hill (2007c) for a discussion of these methods]. The regions can be linked either at the basic heading level or at the aggregate level. Diewert uses a variant on the CPRD method to link at the basic heading level. Before running the regression it is first necessary to convert the prices of all ring countries belonging to the same region into the same base currency using the within-region price indexes generated in stage 1. It does not matter which country is chosen as the base for each region. A CPRD regression is then run over the ring countries for each basic heading with the subtle difference that the country dummies are replaced with region dummies. The estimated coefficients on the region dummies, when exponentiated, provide the transitive between-region price indexes for each basic heading.

To compute price indexes at the regional level at higher levels of aggregation, expenditure vectors are also required for each region. These can be constructed by converting the expenditure vectors of all countries (not just ring countries) in the same region into units of the base country's currency (again the choice of base country for each region does not matter) using the stage 1 between-region price indexes. These expenditure vectors are then summed to obtain the regional expenditure vectors. Any multilateral method can then be used to construct regional price indexes at the aggregate level. Diewert (2004b) does not specify which multilateral method should be used. Diewert's method was adopted (although without the representative product dummies) in ICP 2005. As a matter of consistency the same multilateral method should be used for all regions in stage 1 and to link the regions in stage 2. This is not what happened in ICP 2005 because Africa used the IDB method in stage 1, while all other regions used GEKS. In stage 2, GEKS regional price indexes were used to link five of the six regions (the exception being the CIS region).

As long as all regions use the same list of basic headings, Diewert's method can also be used to construct linked basic heading price indexes for all 146 countries. Rather than computing regional price indexes, it is therefore possible instead to apply a multilateral method such as Geary–Khamis or GEKS directly to the whole world. In practice, this is not advisable for two reasons. First, ICP 2005 has a requirement of fixity in the regional results. For example, this means that the price index for France relative to Germany in the global comparison should be the same as in the Eurostat/OECD comparison. Applying a multilateral method globally in stage 2 would lead to a violation of fixity above the basic heading level. Second, although all regions have the same basic headings, there may be very little overlap from one region to the next in the list of products priced for each basic heading. Hence comparisons between pairs of countries in the same region are likely to be more reliable than comparisons between pairs of countries in different regions. It may be advisable to make use of this fact when constructing the aggregate results.

Hill's (2005) method, by contrast, can be viewed as a generalization of the GEKS method. The GEKS method can be interpreted as the geometric mean of K star comparisons, with each of the K countries placed in turn at the center of the star [see Hill (1997)]. Hill's method starts by computing bilateral links between each pair of regions. Suppose we focus on Eurstat/OECD (with Estonia, Japan, Slovenia, and the United Kingdom as its ring countries) and South America (with Brazil and Chile as its ring countries). Suppose now we select the U.K. and Brazil as the base countries for each region. Depending on the path followed from one region to the other we can obtain eight different estimates of the U.K.–Brazil price index:

	UK–Brazil	
	UK–Chile	× Chile–Brazil
UK–Estonia	× **Estonia–Brazil**	
UK–Estonia	× **Estonia–Chile**	× Chile–Brazil
UK–Japan	× **Japan–Brazil**	
UK–Japan	× **Japan–Chile**	× Chile–Brazil
UK–Slovenia	× **Slovenia–Brazil**	
UK–Slovenia	× **Slovenia–Chile**	× Chile–Brazil

The comparisons in bold above are bilateral comparisons (e.g., using Fisher) between pairs of countries in different regions. The other comparisons are multilateral comparisons (e.g., GEKS) between pairs of countries in the same region. By taking a geometric mean of these eight sets of results, we obtain a price index between Eurostat/OECD and South America that treats all four Eurostat/OECD ring countries symmetrically and both South American ring countries symmetrically. Changing the base country for either region will simply rescale the results. The analogy with GEKS is that each country is used in turn as the link for that region with the other region.

At the end of this process these bilateral indexes between regions must be transitivized using the GEKS formula. Given the method's affinity with GEKS, it should probably be used in conjunction with GEKS comparisons within each region in stage 1. Supposing again that the U.K. and Brazil are the base countries for their respective regions, a price index at the aggregate level between, say, France and Chile is obtained as follows:

$$P_{\mathrm{Fr,Ch}} = P_{\mathrm{Fr,UK}}^{\mathrm{GEKS}} \times P_{\mathrm{UK,Br}}^{\mathrm{Region}} \times P_{\mathrm{Br,Ch}}^{\mathrm{GEKS}},$$

where $P_{\mathrm{Fr,UK}}^{\mathrm{GEKS}}$ denotes a within-region (i.e., Eurostat/OECD) GEKS comparison between France and the U.K., $P_{\mathrm{Br,Ch}}^{\mathrm{GEKS}}$ denotes a within region (i.e., South American) GEKS comparison between Brazil and Chile, and $P_{\mathrm{UK,Br}}^{\mathrm{Region}}$ denotes the overall between-region comparison between Eurostat/OECD and South America with the U.K. and Brazil serving as the base countries for each region.

5. CONCLUSION

ICP 2005 has provided a big stimulus to research in the field of international comparisons. There are some inconsistencies that should be resolved in the next round. The most pressing concern is the effective implementation of the CPRD method to address the representativity bias problem at the basic heading level. Also, ideally all regions should use the same method to construct their basic heading price indexes. At present CIS and Eurostat/OECD use the Jevons-S method, whereas the other regions use the CPRD method. We favor the use of the CPRD method at basic heading level as long as it is properly implemented. Above the basic heading level, all regions use GEKS, with the exception of Africa, which uses the IDB method. At this level, we favor GEKS. GEKS, nevertheless, has its problems. We believe that it may be possible to improve on the GEKS approach by combining it with spanning-tree methods in the next round of ICP. Perhaps the most important innovation of ICP 2005 has been the introduction of a two-stage approach in which each region makes its own comparison in stage 1, and then the regions are linked in stage 2 by means of a comparison between a group of ring countries drawn from the regions. We are supportive of this development. It has, however, required the development of new methods tailored to this two-stage methodology. This is another area that warrants more research. We hope that the current research momentum is maintained into the next round of ICP in 2011.

NOTES

1. Here we focus on prices. Real output measures are derived by deflating nominal output by the appropriate price index.

2. The same distinction exists in temporal indexes such as CPIs and PPIs except that basic headings are called elementary aggregates in the literature on temporal indexes. See Hill (2004).

3. There are 155 basic headings in the ICP. Food and nonalcoholic beverages account for 29 headings; alcoholic beverages, tobacco, and narcotics for 5 headings; clothing and footwear for 5 headings, etc. See Blades (2007).

4. In practice, some countries may be unable to carry out national surveys, in which case the average price for the capital city, or urban areas, may be adjusted to provide an estimate of the national average.

5. See paragraphs 1.133–1.146 of Chapter 1 [Hill (2004)] and paragraphs 20.58–20.86 of Chapter 20 [Diewert (2004a)] of the CPI Manual.

6. This situation is further complicated by the fact that there are products available in one country but not the other. One could imagine computing a reservation price for these products in the country in which they are absent. This would typically lead to products available in country B but not A being placed at the relatively cheap end of the distribution (from country B's perspective) while products available in A but not B would be placed at the other end.

7. The bias here actually acts in the opposite direction to that of the standard Gerschenkron effect.

8. In some cases, either or both of M_{rj-k} and M_{rk-j} may equal zero. In this case, either only one half of formula (1) is used or one reverts to the original Jevons formula for this particular pair of countries.

9. The price index for country 1 is normalized to 1.

10. In the literature on the CPD and in previous ICP reports, betas are generally used to denote countries, although in World Bank (2008) they are used to denote products.

11. We abstract here and in what follows from the fact that when taking a nonlinear transformation of an estimated parameter an adjustment is necessary to prevent bias. See van Garderen and Shah (2002). From our experience in the context of international comparisons this bias is typically small.

12. The estimated error terms were used in ICP 2005 to detect outliers, namely extremely high or low relative prices that are likely to signal errors in the data. Outliers were further investigated and possibly revised or rejected. The method uses tables named after Yuri Dikhanov of the World Bank, who developed the method. See World Bank (2008, p. 198).

13. The existence of a significant correlation between the CPD residuals for certain types of products and level of income can be tested statistically.

14. It should be noted that a complete price table here does not imply that we are in the ideal scenario in which all relevant products are included in the list. Complete here simply means that all countries price every product on the actual as opposed to the ideal list.

15. It is not the case here, however, that Jevons* and CPRD are identical.

16. Indeed, the CPD and CPRD methods can also be used to impute missing prices.

17. When relative prices vary substantially from country to country and representativity is added to the CPD model, the fit may be significantly improved. The multiple correlation coefficient is increased and the standard errors of the estimated PPPs are reduced compared with the simple CPD [see Hill (2007a)].

18. More sophisticated measures of dissimilarity are considered by Diewert (2002).

19. It is important that the bilateral index used satisfies the country reversal test. That is, it must be the case that $P_{jk} = 1/P_{kj}$. Otherwise, directional arrows must be included on each edge, and the problem of selecting the optimal spanning tree becomes rather more complicated. Fortunately, both the economic and axiomatic approaches to index numbers favor formulas that satisfy the country reversal test (i.e., Fisher, Törnqvist, and Walsh). All superlative formulas satisfy this test [see Diewert (1976)].

20. The core here should not be confused with the ring countries used in ICP 2005.

21. Alterman, Diewert, and Feenstra (1999, p. 61) show that if the logarithmic price ratios $\ln(p_{tn}/p_{t-1,n})$ trend linearly with time t and the expenditure shares s_{tn} also trend linearly with time, then the Törnqvist index P^T will satisfy the circularity test exactly. Many economic time series on prices and quantities come reasonably close to satisfying these assumptions. In such cases, in addition to being nearly transitive, Törnqvist will also closely approximate Fisher and Walsh.

22. The 45 countries in the Eurostat/OECD region were actually split into five subregions for purposes of preparing the common lists of products. Each subregion prepared its own separate list of products for each basic heading. The five lists were then merged. If the merged list did not contain enough products that were to be found in two or more different subregions, additional overlapŠ products were added to reinforce the bilateral comparisons between countries in different subregions. However, Eurostat/OECD stopped well short of full regionalization because, for purposes of calculating the elementary PPPs, the prices in the subregions were pooled. Basic heading PPPs were calculated for all five subregions simultaneously. The question of how to link the PPPs for the different subregions therefore did not arise [see Roberts (2006, paragraphs 2.31 to 2.34 and 4.17 to 4.24)].

REFERENCES

Alterman, William F., W. Erwin Diewert, and Robert C. Feenstra (1999) *International Trade Price Indexes and Seasonal Commodities*. Washington, DC: Bureau of Labor Statistics.

Balk, Bert M. (1996) A comparison of ten methods for multilateral international price and volume comparisons. *Journal of Official Statistics* 12, 199–222.

Blades, Derek (2007) GDP and main expenditure aggregates. In *ICP 2003-2006 Handbook*, Chapter 3. Washington, DC: The World Bank.

Cuthbert, Jim and Margaret Cuthbert (1988) On Aggregation Methods of Purchasing Power Parities. OECD Economics Department Working Papers, No. 56. Paris: OECD Publishing.

Deaton, Angus and Alan Heston (2008) Understanding PPPs and PPP-Based National Accounts. NBER Working Paper No. 14499.

Diewert, W. Erwin (1976) Exact and superlative index numbers. *Journal of Econometrics* 4(2), 115–145.

Diewert, W. Erwin (2002) Similarity and Dissimilarity Indexes: An Axiomatic Approach. Discussion Paper 02-10, Department of Economics, University of British Columbia.

Diewert, W. Erwin (2004a) Elementary indices. In *Consumer Price Index Manual: Theory and Practice*, Chapter 20. Geneva: International Labour Organization.

Diewert, W. Erwin (2004b) On the Stochastic Approach to Linking the Regions in the ICP. Discussion Paper 04-16, Department of Economics, University of British Columbia, Vancouver, Canada.

Diewert, W. Erwin (2008) New Methodological Developments for the International Comparison Program. Discussion Paper 08-08, Department of Economics, University of British Columbia, Vancouver, Canada.

Dikhanov, Yuri (1994) Sensitivity of PPP-based income estimates to choice of aggregation procedures. Paper presented at the 23rd General Conference of the International Association for Research in Income and Wealth, St. Andrews, New Brunswick, Canada, August 1994.

Dowrick, Steve and John Quiggin (1997) True measures of GDP and convergence. *American Economic Review* 87(1), 41–64.

Eltetö, Oded and Pal Köves (1964) On a problem of index number computation relating to international comparison. *Statisztikai Szemle* 42, 507–518.

Eurostat (1983) *Comparison in Real Values of the Aggregates of ESA: 1980*. Luxembourg: Statistical Office of the European Communities.

Geary, Roy C. (1958) A note on the comparison of exchange rates and PPPs between countries. *Journal of the Royal Statistical Society*, Series A 121, 97–99.

Gini, Corrado (1931) On the circular test of index numbers. *International Review of Statistics* 9(2), 3–25.

Hill, Robert 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, Robert J. (1999) Comparing price levels across countries using minimum spanning trees. *Review of Economics and Statistics* 81(1), 135–142.

Hill, Robert J. (2000) Measuring substitution bias in international comparisons based on additive purchasing power parity methods. *European Economic Review* 44(1), 145–162.

Hill, Robert J. (2005) A Note on Linking the Regions in the ICP. Mimeo, School of Economics, University of New South Wales.

Hill, Robert J. and T. Peter Hill (2007) Regionalization of the International Comparisons Program: Some Implications for Price Index Construction. Discussion Paper 2007-28, School of Economics, University of New South Wales.

Hill, T. Peter (2004) An introduction to consumer price indices methodology. In *Consumer Price Index Manual: Theory and Practice*, Chapter 1. Geneva: International Labour Organization.

Hill, T. Peter (2007a) Estimation of PPPs for basic headings within regions. In *ICP 2003-2006 Handbook*, Chapter 11. Washington, DC: The World Bank.

Hill, T. Peter (2007b) Ring comparison linking within-region PPPs using between region PPPs. In *ICP 2003-2006 Handbook*, Chapter 14. Washington, DC: The World Bank.

Hill, T.P. (2007c) Linking PPPs and real expenditures for GDP and lower level aggregates. In *ICP 2003-2006 Handbook*, Chapter 15. Washington DC: The World Bank.

Iklé, Doris (1972) A new approach to the index number problem. *Quarterly Journal of Economics* 86, 188–211.

Khamis, Salem H. (1972) A new system of index numbers for national and international purposes. *Journal of the Royal Statistical Society*, Series A 135(1), 96–121.

Rao, D.S. Prasada (2004) The Country-Product-Dummy Method: A Stochastic Approach to the Computation of Purchasing Power Parities in the ICP. Paper presented at the SSHRC Conference on Index Numbers and Productivity Measurement, Vancouver, July 2004.

Roberts, David (2006) *Eurostat–OECD Methodological Manual on Purchasing Power Parities (PPPs)*. Paris: OECD.

Sergeev, Sergey (2003) Equi-representativity and Some Modifications of the EKS Method as the Basic Heading Level. Paper presented at the Joint Consultation on the European Comparison Programme, ECE, Geneva, 31 March–2 April 2003.

Summers, Robert (1973) International price comparisons based upon incomplete data. *Review of Income and Wealth* 19(1), 1–16.

Szulc, B.J. (1964) Indices for Multiregional Comparisons. *Przeglad Statystyczny* 3 [*Statistical Review* 3], 239–254.

van Garderen, Kees Jan and Chandra Shah (2002) Exact interpretation of dummy variables in semilogarithmic equations. *Econometrics Journal* 5, 149–159.

Wilson, Robert J. (1985) *Introduction to Graph Theory*, 3rd ed. New York: Longman.

World Bank (2008) *Global Purchasing Power Parities and Real Expenditures, 2005 International Comparison Program*. Washington, DC: The World Bank.