Chapter 4

Computation of Basic Heading Purchasing Power Parities (PPPs) for Comparisons within and between Regions

D.S. Prasada Rao
## Contents

1. Introduction .................................................................................................................................................. 3

2. Features of Data for the Computation of PPPs at the BH level................................................................. 4

3. Aggregation Methods for Computing PPPs at BH Level: Analytical Requirements ......................... 9

4. Jevons Index and the Gini-Elteto-Koves-Szulc (GEKS) Method for PPPs at the BH Level .... 11

5. The Country-Product-Dummy (CPD) Method for the Computation of BH PPPs ......................... 16

6. Linking Regions at the Basic Heading level in 2005 ICP and Looking Forward to 2011 ICP .. 24

7. Conclusions .................................................................................................................................................. 28

References ......................................................................................................................................................... 29
Computation of Basic Heading Purchasing Power Parities (PPPs) for Comparisons within and between Regions

1. Introduction

This chapter is the first of a set of three chapters of this book devoted to the description of the aggregation methods used at various stages in the compilation of purchasing power parities of currencies of countries within regions and for comparisons between regions. A schematic diagram showing various stages involved in the computation of PPPs is given in Figure 1.3 in Chapter 1. The main objectives of this chapter are to provide a detailed description of the methods used in the computation of PPPs at the basic heading (BH) level at the regional level and to describe how BH level PPPs are compiled for the purpose of making global comparisons of prices. The next chapter by Erwin Diewert focuses on the methods for aggregation of price and quantity data in the computation of PPPs for higher level aggregates such as Consumption, Investment, and, finally, the GDP at the regional level. Chapter 6 by Erwin Diewert examines the problem of linking regional comparisons above the BH level to make global comparisons.

The literature on aggregation methods for international comparisons has traditionally focused on the aggregation at levels above the basic heading level (see Balk, 2008; and Rao, 2009 for excellent reviews of such methods) and relatively little weight is placed on the problem of constructing PPPs at the BH level. In a recent paper, Hill and Hill (2009) stress the importance of PPPs at the BH level through the following quote:

“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 basic heading price indexes provide 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.” (Hill and Hill, 2009, pp. 192-193)

The importance of reliable PPPs at the BH level has been reiterated after the release of the 2005 ICP results (World Bank, 2008) which have shown substantial revisions to the global GDP and to the real incomes of major economies like China and India. For example, Deaton and Heston (2008) examine possible reasons for such revisions and conclude that one of the underlying factors is the price data used for compilation of PPPs at the BH level.

Reflecting the concern to improve the quality and reliability of PPPs at the BH level, the material in this chapter is designed to provide an appreciation of the core issues surrounding the compilation of PPPs at the BH level including a description of the characteristics of the price data as well as the characteristics of the

---

* The chapter is prepared for inclusion in MEASURING THE SIZE OF THE WORLD ECONOMY: A Framework, Methodology and Results from the International Comparison Program (ICP) (eds. D.S. Prasada Rao and Fred Vogel). The author wishes to acknowledge comments from Erwin Diewert, Robert Hill, Sergey Sergeev and Fred Vogel on the material covered in this chapter. Additional thanks are due to Sergey for providing data and results for the numerical example included in this chapter.

1 PPPs compiled at the regional level are based exclusively on data for the countries included in the region. These are expected to satisfy the usual properties of transitivity and base invariance only for the countries in the region. In contrast global comparisons make use of data from all the countries included in the global comparison and PPPs at the level satisfy transitivity across countries from different regions.
products priced in different countries. The chapter seeks to illustrate how the aggregation methods used in the ICP are designed to take full account of information on such characteristics. Reflecting the forward looking nature of the ICP Book, the current chapter not only reviews the methods used in the 2005 ICP but also describes the new approaches and methods being considered for the 2011 ICP. A significant new approach in the 2011 ICP concerns the use of a list of core products for the purpose of linking regions. While there is a continuing discussion on how to calibrate the aggregation methods to best utilize prices on core products, consensus has emerged on the method to be used for linking PPPs at the BH level.\(^2\)

The chapter draws material from a number of sources including: (i) the extensive discussion of the issues surrounding aggregation at the BH level presented in the ICP Handbook 2003-2007 (World Bank, 2006); (ii) the Final Report of the 2005 ICP from the Global Office (World Bank, 2008); and (iii) from various published and unpublished papers and reports on the subject.

The chapter is organized as follows. Section 4.2 focuses on features and aspects specific to comparisons at the BH level. Details of the nature of price data, product characteristics and the features of price quotations that have direct bearing on the aggregation methods used are all discussed in this section. Section 4.3 briefly discusses the analytical requirements for the aggregation methods used at the BH level. Sections 4.4 and 4.4 are devoted to the description of aggregation methods used in the compilation of PPPs at the regional level. The Jevons binary price indexes and the Gini-Elteto-Koves-Szulc method of constructing transitive parities are the subject matter for Section 4.4. Variants of the Jevons-GEKS indices that are designed to make use of information on representativity and importance of products\(^3\) are also discussed in this section. Section 4.5 is devoted to the description of the country-product-dummy (CPD) method and its variants used in the computation of BH PPPs\(^4\). Section 4.6 focuses on the problem of linking the regional BH PPPs and describes the methodology used in the 2005 ICP based on ring countries and ring product list as well as the new approach based on the use of core products proposed for the 2011 ICP. Section 4.7 provides some concluding remarks.

2. Features of Data for the Computation of PPPs at the BH level

As a starting point it is useful to note the main feature of the BH level data and what distinguishes the basic heading level from higher levels of aggregation. The features discussed below have a direct bearing on the methodology used in aggregating data. The number of products in the basic heading priced and the relative overlap across the countries can lead to differences in results from different methods. If all the products are priced in all the countries and if all the products are treated as equally important then the two major approaches used within the ICP, the Jevons index with GEKS and the CPD methods, will lead to identical PPPs thus eliminating the need to choose between methods. In addition it is important to understand the concept of basic heading, and the concepts of representativity and importance before identifying an appropriate index number method.

\(^2\) At its meeting in April, 2011, the Technical Advisory Committee has recommended procedures for linking PPPs across regions at the BH level.
\(^3\) Representativeness of products was briefly discussed in Chapter 1. It will be further elaborated in this chapter.
\(^4\) The CPD and the CPRD (country-product-representativity-dummy) methods were recommended for use in 2005 ICP at the BH level.
Basic Heading

Within the pyramid approach to price comparisons, as illustrated in Figure 1.1 in Chapter 1, price comparisons begin with the identification of products in the regional list of items for which price data are collected through price surveys. These products are grouped into different basic headings. As a working definition, basic heading is the lowest level of aggregation for which expenditure share weight information is available. An immediate implication is that in the ICP, no quantity data are available for the products priced in the surveys. Thus price comparisons at the BH level are derived as aggregates of item level price ratios and, therefore, BH PPPs are similar to the elementary indices compiled as a part of the compilation of the consumer price index (CPI). In addition to being the smallest aggregate with expenditure share weights, a general consideration is that groups of similar products defined within a general product classification will constitute basic headings. Similarity of products within the basic heading is supposed to ensure the similarity in the magnitudes of price relatives for products included in the BH. However, in the ICP the composition of basic headings is necessarily dictated by the detailed classification used in the national accounts and the availability of national accounts’ expenditure share weights for a desired grouping of products. Therefore, it is possible that basic headings used in ICP price comparisons may not always include products which may exhibit similar price movements. For example, the Clothing and Footwear aggregate consists of the following five basic headings: (i) Clothing materials; (ii) Garments (men’s, women’s and children’s garments); (iii) Cleaning, repair and hire of clothing; (iv) Shoes and other footwear; and (v) Repair and hire of footwear. In this example, it is clear that similar products are grouped together. However, there is considerable heterogeneity within groups. In the Garments group includes men’s, women’s and children’s clothing and, further, there would be further heterogeneity within men’s garments which include shirts, trousers, suits as well as undergarments. It is difficult to see if all the products within the Garment’s group is likely to exhibit similar price relatives or price movements.

From the perspective of computing a PPP for a given basic heading, the most ideal situation is where item specific weights in the form of quantities or expenditure share weights are available. In the absence of such information, several characteristics of the products included in the basic heading are taken into account in computing BH PPPs. It is useful to review these characteristics as they play a critical role in the choice of an appropriate formula for the computation of BH PPPs.

Product Characteristics

In the 2005 ICP a new approach based on the structural product descriptions (SPDs) was introduced. The SPDs provide a structured and systematic way to describe the price determining characteristics for products households purchase to ensure that prices collected in different countries refer to the same products thus ensuring comparability. The SPDs were used to prepare product specifications which were provided to the price collectors. For example, in the Asia-Pacific region 20 different varieties of rice were defined by the rice SPD for pricing purposes (ADB, 2007). The introduction of SPDs ensured comparability of products priced in different countries. However, it is widely recognized that a product which is comparable in its characteristics across countries may not be representative of consumption in all the countries where the

---

5 The survey framework for collection of price data is described in Chapter 7 by Fred Vogel.
6 Lack of quantity or expenditure share data makes it impossible to use standard index number formulae like the Laspeyres, Paasche, Fisher and Tornqvist index number formulae.
7 For a description of elementary indices and the methods used, see the ILO-ECE (2007) manual for the compilation of the CPI.
8 In the case of temporal movements in prices, products within a basic heading are expected to show similar price changes over time.
product is priced. In general, there is a trade-off between comparability and representativity in identifying products for inclusion in the basic headings and price surveys.

Representativity

The concept of representativity was developed as a part of the Eurostat-OECD PPP program as a way of accounting for the possibility that some of the items in the product lists may not be representative of consumption in some of the participating countries. It makes intuitive sense that prices of representative products should have greater influence on the PPP for the basic heading and converse should be true with unrepresentative products. The actual implementation requires a formal process for identifying representative products.

Representativity of a product is determined on two main considerations.

- The first consideration is that a particular item is representative in a country for a particular basic heading if the particular product has a significant market share as reflected by the expenditure share or by the volume of sales associated with a particular item. This notion of representativity is consistent with the approach one would take if quantity or expenditure data were available at the item level. In the presence of such information, aggregation of price ratios would be weighted proportional to the expenditure shares. However as expenditure share data are not available at the item level, labeling a product as representative has to be left to the price statistician or local experts.

- The second consideration is based on the notion that representative products have lower price levels than unrepresentative products and not accounting for such differences in aggregating price data could result in biased PPPs. If products could be labeled as representative or unrepresentative in each of the countries, such information could be incorporated into the aggregation process. However it is difficult to decide a priori whether a particular product is representative on the basis of whether the price relative of the product in two countries is typical of the products in the basic heading. A possible approach could be based on the fact that a cheap product would be purchased in large quantities and therefore would be popular.

An important reason for distinguishing between representative and unrepresentative products is that relative prices of these products would be different. Products purchased in large quantities are associated with lower prices relative to other products in the basic heading and unrepresentative products may have higher relative prices. On the basis of this, a product may be considered unrepresentative if its relative price is high and therefore may not be representative of the relative prices of products which are representative and included in the basic heading. This notion of a representative commodity underpins the use of the country-product-representativity-dummy (CPRD) and the starred-EKS methods used in international comparisons.

Though the conceptual basis for identifying representative products within a basic heading is sound and appealing, in practice the national and regional ICP coordinators have no real objective measures to decide whether a particular product is representative or not. In particular, the notion of whether it is representative

---

9 For more details of the OECD-Eurostat treatment of representativity see the paper by Roberts (2009) where a comprehensive account of the methodology used by OECD and Eurostat is provided.

10 Consider a hypothetical example where country A prices only representative products and country B prices only unrepresentative products. In this case the PPP for country B relative to country A would overstate the price levels in these two countries.

11 This may not hold for all products. For example, beef is not representative of meat consumption in India but it is not relatively more expensive than other meats like lamb or chicken. In fact beef and pork are generally cheaper as they are not commonly consumed.
on the basis of relative prices is a difficult concept to implement. These issues underscored the failure of the national and regional coordinators to meaningfully identify representative products. In the Asian and African regions, there was a lot of confusion in the identification of representative products to the extent that including such information induced biases into the estimated PPPs at the BH level. Thus the information collected on representativeness of products within basic headings was discarded at the stage of computing PPPs for basic headings. The OECD-Eurostat and CIS regions are the only regions which made use of information on representativeness.

Importance of Products

In view of the difficulties encountered during the 2005 ICP, it was decided that the concept of representativeness would be replaced by an indication whether a product is important among the list of products within a basic heading. Aggregation methods leading to BH PPPs should explicitly account for the importance of a particular item in a particular country with unimportant products accorded a smaller weight. The notion of importance is supposed to reflect if either the sales or the expenditure share are significant enough to warrant the label that the product is important. The national statisticians have a reasonable understanding of the importance as reflected by the sales or shares. The notion of importance of the product is not related in any way to the relative price of the product and, therefore, it is a weak alternative to the use of actual quantity data along with price data in the computation of PPPs at the BH level. It was generally agreed that an important product will be accorded a weight of 3 which is three times the weight attached to the remaining products which are considered unimportant.\(^\text{12}\)

Price Data – National Average Prices

The price concept relevant for the ICP is the annual national average price of the item priced in each of the countries. In the 2005 ICP, national average prices are computed as arithmetic averages of a large number of price quotations for each of the items in the product lists.\(^\text{13}\) The average price should ideally cover the whole country representing the rural and urban regions as well as different geographical regions in the case of larger countries. The following data are supplied by the countries to the regional coordinator of the ICP:

- National annual average price;
- Number of price quotations used in the computation of the average price;
- The standard deviation of the price quotations used in computing the national average price.

In this case, the reliability of a given national average price can be computed as the standard error associated with the national average price.\(^\text{14}\)

The PPPs computed at the BH level must make use of all the information available associated with the price data (national average prices) from the countries participating in the comparisons. In particular, it is statistically desirable to accord lower weights to prices that have higher standard errors associated. Though this is a desirable approach, standard errors of the national average prices have so far not been incorporated into the PPP computation.

---

\(^{12}\) This is the recommendation made at the last meeting of the Technical Advisory Group in April, 2011. These weights will be used in conjunction with the CPD method. Details of the procedure are given in Section 4.4.

\(^{13}\) For a discussion on the determination of the number of quotations and other related issues, see Chapter 7 by Fred Vogel on the Survey Framework for the ICP.

\(^{14}\) If \(\sigma\) is the standard deviation associated with a given price (average) and if \(n\) is the number of price quotations used, then \(\sigma / \sqrt{n}\) is the standard error associated with the given average price.
Basic Data for the Computation of BH PPPs

This subsection establishes the notation used throughout this chapter and describes the nature of price data used in the compilation of PPPs at the BH level. For the purpose of this exposition, without loss of generality focus is on a single region and a selected basic heading. Let $N$ represent the number of commodities included in a given BH and $C$ represent the number of countries included in a given region. Let $p^c_i$ denote the price of $i$-th commodity in country $c$ ($i=1,2,\ldots,N; c=1,2,\ldots,C$) where the price is assumed to be strictly positive. The problem of determining the number of items and the actual items for inclusion in the BH is dealt with in the Chapter 7 by Fred Vogel on Survey Methodology for the ICP. In practice, the following two scenarios are possible:

- all the items are priced in all the countries leading to a complete tableau or matrix of prices;
- not all items are priced in all the countries leading to an incomplete tableau of prices; and
- some items may be priced in only one of the countries in the region.

Table 1 gives two examples of incomplete tableau.

Table 1: Two Examples of Incomplete Tableau

<table>
<thead>
<tr>
<th>Items in BH</th>
<th>Prices in National Currency Units</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TABLEAU I</td>
</tr>
<tr>
<td></td>
<td>Countries</td>
</tr>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Long grain</td>
<td>10</td>
</tr>
<tr>
<td>Medium grain</td>
<td>12</td>
</tr>
<tr>
<td>Short grain</td>
<td>-</td>
</tr>
<tr>
<td>Low quality</td>
<td>4</td>
</tr>
<tr>
<td>Imported</td>
<td>25</td>
</tr>
</tbody>
</table>

In Tableau I, price data are missing in a number of cells but it would be possible to compute PPPs for the four countries in the comparison as the data exhibit overlaps between countries. Note that the fourth item, Low Quality rice, is priced only in country A and therefore cannot be used in PPP computation. Therefore at the BH level in order to influence the PPP computations an item must be priced in at least two countries. Tableau II is an interesting case where the Long Grain and Imported rice items are priced only in countries A and B. In contrast, only the Medium Grain and Short Grain are priced in countries C and D. While it is possible to compare prices across A and B and between C and D, it is not possible make price comparisons between countries A and C or A and D. Similarly B cannot be compared with either C or D. In this case the price tableau can be reduced into two blocks, one for the two countries A and B and another block consisting of countries C and D with nothing in common between the blocks.

15 Generally $N$, number of items, would vary with the basic heading and therefore ideally a subscript needs to be added. Similarly, the number of countries varies across different regions and therefore $C$ needs to have regional descriptor. As the focus is on a single region and on single basic heading, for expositional purposes, subscripts for the basic heading and the regions are dropped.

16 This case is same as the case where all the countries in the region price the same subset of items which implies that the remaining items are not priced in any of the countries and can therefore be dropped from the computations.
Several important conclusions can be drawn from the examples in Table 4.1.

- First, when the price tableau is incomplete, it is important that the price data collected should be connected and price comparisons between all the countries involved can be made if and only if the price tableau is connected or irreducible. The price tableau is said to be connected if the price data are such that it is not possible to group the countries into two groups such that no item priced by any country in one group is priced by any other country in the second group. In Tableau II above, it is possible to make countries into two groups (A and B; and C and D) such that no rice item priced in the first group is priced by any country in the second group and vice versa. In such cases the process of price comparisons breaks down.\(^\text{17}\)

- Second, if a product is priced in only one country its price will have no influence on the PPPs computed.

- Third, when price tableau is incomplete, the quality of price comparisons depends upon the strength of interconnections and overlaps in the priced items across different countries. If the overlaps are strong then one can make reliable price comparisons across different countries. A corollary of this observation is that countries within a particular region should strive to price as many items in the basic heading as possible\(^\text{18}\). Chapter 7 by Fred Vogel on the Survey Framework for the ICP provides guidelines regarding the minimum numbers of items to be priced within each basic heading.

3. Aggregation Methods for Computing PPPs at BH Level: Analytical Requirements

In order to identify the properties expected of aggregation methods for computing PPPs at the BH level, it is useful to recall a working definition of PPP and then apply it to the special case of a single product. A purchasing power parity (PPP) between currencies of countries A and B may be defined as the number of currency units of country A that have the same purchasing power of one unit of currency of country B defined with respect to a designed product or set of products such as a basic heading.

**PPPs for individual products**

Suppose \( p^j_i \) and \( p^k_i \) are respectively the prices of product \( i \) in countries \( j \) and \( k \), then PPP for country \( k \) with respect to country \( j \) is given by:

\[
PPP_{jk} = \frac{p^k_i}{p^j_i}
\]  

(4.1)

Obviously, PPP defined in (4.1) depends upon the particular product selected. For a selected commodity \( i \), the following transitivity property can be noticed. For any three countries \( j, k \) and \( m \) it is easy to see that:

\[
PPP_{jk} = \frac{p^k_i}{p^j_i} = \frac{p^k_i}{p^m_i} \cdot \frac{p^m_i}{p^j_i} = PPP_{jm} \cdot PPP_{mk}
\]  

(4.2)

\(^{17}\) In the case where the price data are not connected, then even the idea of using a spatial chain of countries will fail as there is no way to connect countries A and B with countries C and D in the second block.

\(^{18}\) As price data collection is resource intensive, a balance should be struck between the cost of collection and the need to price as many items as possible to strengthen the price comparisons.
Equation (4.2) shows that PPP between countries $j$ and $k$ is equal to the indirect PPP comparison derived through a third country, $m$. Equation (4.2) guarantees a level of internal consistency required for international comparisons. This property is known as the transitivity property and equation (4.2) shows that when PPPs are based on prices of a single product, then transitivity is automatically satisfied.

**Transitivity for PPPs in Multilateral and Multiproduct comparisons**

Multilateral PPPs, represented by the matrix of PPP comparisons between all pairs of countries, based on price data on more than one item (that is in the presence of multiple products) is said to be transitive if for any three countries in the group, such as $j$, $k$ and $m$, the direct PPP for country $k$ with respect to country $j$ is equal to the indirect PPP derived through the use of the third country, $m$.

$$PPP_{jk} = PPP_{jm} \cdot PPP_{mk} = \frac{PPP_{mk}}{PPP_{mj}}$$

(4.3)

The last part of equation (4.3) requires the assumption that $PPP_{jm}$ is the reciprocal of $PPP_{mj}$.

As the PPPs based on a single product defined in (4.1) are automatically transitive and further that PPP's based on price data for multiple items in a basic heading would require some sort of averaging of item level PPPs, it is necessary to consider only those methods that retain the property of transitivity as stated in (4.3). Unless stated otherwise, all the procedures considered in the reminder of this section satisfy transitivity.

**Base Invariance of PPPs for Multilateral Comparisons**

In addition to the transitivity requirement stated in (4.3), it is important that all the countries involved in the price comparisons are treated symmetrically and that no country is accorded a special status. This condition is particularly relevant as it is possible to generate transitive PPPs using a “star” method where a single country like the United States is at the centre and all countries are compared only through the star country (see Figure 1.2 in Chapter 1 for an illustration). While such comparisons based on a selected star are transitive, the comparisons are sensitive to the start country selected. For example, relative comparisons and PPPs between two countries, say $j$ and $k$, would be different when two different countries, say the United States and Germany, are used as star countries. Again, unless stated all the methods discussed in the ensuing sections produce PPPs that are base invariant.

Traditionally there have been two main approaches to aggregation at the BH level. The first approach is based on the Jevons index used in the computation of elementary price index numbers and the Gini (1924), Elteto-Koves (1964) and Szulc (1964) which is referred to as the GEKS method. An alternative approach developed originally by Summers (1973) makes use of a regression model known as the country-product-dummy (CPD) model as a way of filling or imputing missing price data. However, it was also used as a method of aggregation below the BH level in the earlier rounds of ICP conducted by Kravis and his associates (Kravis, Heston and Summers, 1982). In the recent years the model has received attention through the work of Rao (1990, 1995, 2004, 2005, 2009), Sergeev (2002, 2003), Diewert (2004, 2005, 2010), Hill (2007), Rao and Timmer (2004) and Hill and Timmer (2006). Though no specific references are given further, the following material draws heavily from Hill (2007) which is Chapter 11 for the ICP Handbook for the 2005 ICP and on Diewert (2010). The next two sections of this chapter describe the two major approaches to the compilation of PPPs at the BH level.
4. Jevons Index and the Gini-Elteto-Koves-Szulc (GEKS) Method for PPPs at the Basic Heading Level

The section describes the methodology used by the Eurostat since the 1980s and the current OECD-Eurostat methodology (see Roberts, 2009 for detailed description of the OECD-Eurostat PPP Program) for the purpose of aggregating item level price data to compute BH level PPPs. The basic element of the Eurostat approach is the Jevons index which is the main index number formula used in the computation of elementary price indices in the compilation of the consumer price index. A comprehensive discussion of elementary indices and the properties of the Jevons index can be found in Diewert (2004). The Jevons index by itself does not yield transitive comparisons except in the special case where all the countries price all the products in the basic heading. The Jevons index is suitably transformed using the GEKS approach and used in the computation of BH PPPs. As the Eurostat-OECD program collects reliable information on representativity of different items in the BH in different countries, the Jevons-GEKS method is further modified to account for the additional information on representativity. A variation of the Eurostat-OECD approach proposed by Sergeev (2003). Thus the following scenarios are considered below.

(i) First, all items are priced in all the countries (a complete tableau) with no weights attached to the items reflecting representativity or importance. In this case the standard Jevons index is used.

(ii) Second scenario refers to an incomplete price tableau where not all the items are priced in all the countries but all the items are treated with equal weight. Here the Jevons-GEKS index is used in deriving transitive comparisons.

(iii) Third scenario refers to the most general case where the price tableau is incomplete and at the same time distinction is made between representative and unrepresentative commodities. As representative products are marked with asterisk (*), the method used denoted as the Jevons-GEKS(*) index. A variation of this method proposed by Sergeev (2003) is referred to as the Jevons-GEKS(*)-S index.

**Jevons Index – Complete Price Tableau without weights**

In the simplest case where all the $N$ items are priced in all the countries and treating all the items as equally important in the absence of any implicit weights, the PPPs at a given BH level can be computed using:

$$PPP_{jk}^{JEVONS} = \prod_{i=1}^{N} \left[ \frac{p_{ik}}{p_{ij}} \right]^{1/N} \text{ for all } j, k = 1, 2, ..., C$$

(4.4)

The index in (4.4) is a simple geometric mean of all the price relatives for countries $j$ and $k$ for all the commodities in the BH. This formula is referred to as the Jevons index in the index number literature (see Diewert, 2004 for more details).

It is easy to check that the Jevons index in (4.4) results in PPPs that are transitive and base invariant. It is also useful to note here that in this case where all the $N$ items are priced in all the countries, the CPD method discussed in the next section produced PPPs identical to those based on the Jevons index.

---

**Jevons-GEKS Index – Incomplete Tableau without weights**

Consider the case where not all commodities are priced in all the countries. Let \( N_j \) be the number of commodities, out of \( N \), priced in country \( j \). In addition, suppose that all the price data are connected so that price comparisons are feasible. Note that any binary comparison between countries \( j \) and \( k \) can be made on the basis of overlapping price data consisting of the items that are commonly priced. If a commodity is not priced in one of the two countries that commodity cannot be included in the PPP computation. Let \( N_{jk} \) represent the set and number of commodities in the basic heading that are commonly priced in countries \( j \) and \( k \). Then the PPP for a binary comparison between \( j \) and \( k \) is given by:

\[
PPP_{jk}^{JEVONS} = \prod_{i \in N_{jk}} \left( \frac{p_{j}^i}{p_{j}^j} \right)^{1/N_{jk}} \tag{4.5}
\]

The binary PPP for countries \( j \) and \( k \) based on the commonly priced items given in (4.5) does not satisfy the transitivity property. The GEKS procedure is a technique that generates transitive multilateral indices (PPPs) which may be denoted by \( PPP_{jk}^{EKS} \). Details of the GEKS procedure can be found in Balk (2009) and Rao (2009). The GEKS based PPPs are given by:

\[
PPP_{jk}^{Jevons-GEKS} = \prod_{i \in N_{jk}} \left[ PPP_{ji}^{Jevons} \cdot PPP_{ik}^{Jevons} \right]^{1/C} = \prod_{i \in N_{jk}} \left[ \prod_{i \in N_{jk}} \left( \frac{p_{j}^i}{p_{j}^j} \right)^{1/N_{jk}} \cdot \prod_{i \in N_{jk}} \left( \frac{p_{k}^i}{p_{k}^j} \right)^{1/N_{jk}} \right]^{1/C} \tag{4.6}
\]

It is easy to check that the Jevons-GEKS PPPs in (4.6) satisfy transitivity. Rao and Maddison (2003) provide an intuitive interpretation of the PPPs based on the GEKS method which also establishes its base invariance property. If the price tableau is complete then it is possible to show that the use of Jevons-GEKS indices in (4.6) will lead to PPPs identical to those in equation (4.4).

**Jevons-GEKS(∗) Index – Incomplete Tableau with Asterisks for Representative Items**

Consider the case where products are labeled as “representative” or “unrepresentative” in different countries. Representative commodities are marked with an asterisk (∗). In this case, the OECD-Eurostat comparisons are based on a modified Jevons-GEKS method which is also known as Jevons-GEKS∗ method. The modification is driven by the fact that for any given pair of countries \( j \) and \( k \): (i) there may be a set of products which are priced in both countries which are considered representative; (ii) a set of products priced that are representative in country \( j \) but not representative in country \( k \); (iii) a set of products priced that are representative in country \( k \) but not in country \( j \); and, finally, (iv) products priced in both countries but not representative in either of the countries. This approach can also be used when “representativeness” is replaced by “importance” of commodities. The following notation is used below:

Let \( N_{jk} \) represent the number of products that are representative in either country \( j \) or in country \( k \) and for which price data are reported in both countries \( j \) and \( k \). \( N_{jk} \) will be generally smaller than the total number of commodities \( N \) in the basic heading.

Let \( N_{jk}^{R} \) represent the set and number of products that are representative in country \( j \) which are also priced in country \( k \). Note that they may not all be representative in country \( k \).
Let \( N_{kj}^R \) represent the set and number of products that are representative in country \( k \) which are also priced in country \( j \). Note that they may not all be representative in country \( j \).

Then the PPP for a binary comparison between \( j \) and \( k \) based on only representative or asterisked commodities in country \( j \), denoted by \( \text{PPP}_{jk}^{\text{Jevons}(\ast)} \), is given by:

\[
\text{PPP}_{jk}^{\text{Jevons}(\ast)} = \prod_{i \in N_{kj}^R} \left[ \frac{p_{ik}^j}{p_i} \right]^{\frac{1}{N_{kj}^R}} \quad (4.7)
\]

However, an equally meaningful PPP measure can be defined using commodities that are representative in country \( k \) which are also priced in \( j \), denoted by \( \text{PPP}_{jk}^{\text{Jevons}(k\ast)} \) and given by:

\[
\text{PPP}_{jk}^{\text{Jevons}(k\ast)} = \prod_{i \in N_{kj}^R} \left[ \frac{p_{ik}^j}{p_i^j} \right]^{\frac{1}{N_{kj}^R}} \quad (4.8)
\]

In a statistical or analytical perspective, the two PPP measures given in equations (4.7) and (4.8) are equally desirable as each respectively makes use of the representative products priced in the country which are also priced in the other country. Therefore an asterisk (*) based Jevons index of PPP between \( j \) and \( k \) may be defined using a geometric average of the two PPPs in (4.7) and (4.8). The “representative” or (*) based Jevons index, denoted by \( \text{PPP}_{jk}^{\text{Jevons}(\ast)} \) which is given by:

\[
\text{PPP}_{jk}^{\text{Jevons}(\ast)} = \left[ \text{PPP}_{jk}^{\text{Jevons}(\ast)} \cdot \text{PPP}_{jk}^{\text{Jevons}(k\ast)} \right]^{1/2} = \left[ \prod_{i \in N_{kj}^R} \left[ \frac{p_{ik}^j}{p_i} \right]^{\frac{1}{N_{kj}^R}} \prod_{i \in N_{kj}^R} \left[ \frac{p_{ik}^j}{p_i^j} \right]^{\frac{1}{N_{kj}^R}} \right]^{1/2} \quad (4.8)
\]

The new index is the geometric mean of the indices in (4.7) and (4.8).

Given that \( \text{PPP}_{jk}^{\text{Jevons}(\ast)} \) makes use of only information regarding countries \( j \) and \( k \) from the price tableau, the resulting indices are not transitive even when the price tableau is complete. Therefore, it is necessary to make use of the GEKS procedure which results in a transitive PPPs incorporating “representativity” information at the same time. This index is Jevons-GEKS(*) index. Basically the \( \text{PPP}_{jk}^{\text{Jevons}(\ast)} \)'s are used along with the GEKS approach leading to \( \text{PPP}_{jk}^{\text{Jevons-GEKS}(\ast)} \) for all \( j \) and \( k \). These are given by:

\[
\text{PPP}_{jk}^{\text{Jevons-GEKS}(\ast)} = \prod_{i=1}^{C} \left[ \text{PPP}_{jk}^{\text{Jevons}(\ast)} \cdot \text{PPP}_{jk}^{\text{Jevons}(\ast)} \right]^{1/C} \quad (4.9)
\]

The \( \text{PPP}_{jk}^{\text{GEKS}(\ast)} \)'s given in (4.9) are transitive and base invariant.

---

20 Hill and Hill (2009) uses a more complex notation in presenting these methods.

21 Hill and Hill (2009) interpret this formula as a variant of the standard Tornqvist index which uses geometric mean of price relatives with average budget shares in two periods as the weight.
These were in use at the OECD-Eurostat until the recently proposed modification of the Jevons-GEKS(*) due to Sergeev (2003). This method is referred to as Jevons-GEKS*(S) method. The Sergeev (2003) approach explicitly recognizes and provides additional weights to those items that are representative in both countries and also priced in both countries.

**Jevons-GEKS**(S) Index – Incomplete Tableau with Asterisks for Representative Items with Differential Weights

The method is similar to that of GEKS* method. So a similar notation may be used.

Let \( N^R_{jk} \) represent the set and number of products that are representative in country \( j \) which are also priced in country \( k \) but not representative in \( k \).

Let \( N^R_{kj} \) represent the set and number of products that are representative in country \( k \) which are also priced in country \( j \) but not representative in country \( j \).

Let \( N^W_{jk} \) be the number of commodities that are priced in both countries and also representative in both countries.

Then the PPP for a binary comparison between \( j \) and \( k \) based on only representative commodities in country \( j \) which are not representative in \( k \), denoted by \( PPP^\text{Jevons}(j\rightarrow k) \) and given by:

\[
PPP^\text{Jevons}(j\rightarrow k) = \prod_{i \in N^W_{jk}} \left[ \frac{p^k_i}{p^j_i} \right]^{\frac{1}{N^W_{jk}}} \tag{4.10}
\]

However, an equally meaningful PPP measure can be defined using commodities that are representative in country \( k \) which are also priced in \( j \) but not representative in \( j \), which is denoted by \( PPP^\text{Jevons}(k\rightarrow j) \) and given by:

\[
PPP^\text{Jevons}(k\rightarrow j) = \prod_{i \in N^W_{kj}} \left[ \frac{p^k_i}{p^j_i} \right]^{\frac{1}{N^W_{kj}}} \tag{4.11}
\]

The Sergeev (2003) method uses a third index which is based purely on those commodities which are priced in both countries and are representative in both countries. The third index is given by:

\[
PPP^\text{Jevons}^{(*)} = \prod_{i \in N^W_{ij}} \left[ \frac{p^k_i}{p^j_i} \right]^{\frac{1}{N^W_{ij}}} \tag{4.12}
\]

In a statistical or analytical perspective, there is no way of choosing between the three PPP measures given in equations (4.10), (4.11) and (4.12) as each respectively makes use of the representative products priced in the country which are also priced in the other country. Therefore an asterisk (*) based index of PPP between \( j \) and \( k \) may be defined using a geometric average of the three PPPs in (4.10), (4.11) and (4.12) which is weighted by the number of products in different groups is given below. In particular, the Sergeev (2003)
method gives double the weight for the index based on representative products in both countries. The resulting modified “representative” product based PPP, denoted by $PPP_{jk}^{Jevons(S)}$ which is given by:

$$PPP_{jk}^{Jevons(S)} = \left[ PPP_{jk}^{Jevons(\ast \ast)} \right]^{w_1} \left[ PPP_{jk}^{Jevons(j \rightarrow \ast)} \right]^{w_2} \left[ PPP_{jk}^{Jevons(k \rightarrow \ast)} \right]^{w_3} \quad (4.13)$$

where $w_1 = \frac{2 \cdot N_{jk}^{**}}{2 \cdot N_{jk}^{**} + N_{jk}^R + N_{kj}^*}$; $w_2 = w_3 = 0.5 \left[ \frac{N_{jk}^R + N_{kji}^*}{2 \cdot N_{jk}^{**} + N_{jk}^R + N_{kj}^*} \right]$

A simple example may be useful in understanding the weighting scheme used here. Suppose there are 12 items commonly priced in the two countries $j$ and $k$. Let items 1 to 7 be the products that are representative in $j$ but not in $k$; products 8 to 10 are representative in both countries; and that products 11 and 12 are representative in $k$ but not in $j$. Then

$$N_{jk}^R = 7; N_{jk}^{**} = 3; N_{kj}^R = 2$$

and in this case: $w_1 = \frac{2 \cdot 3}{2 \cdot 3 + 7 + 2} = \frac{6}{15} = 0.4; w_2 = w_3 = 0.3$.

Now to complete the procedure by generating transitive indices, it is necessary to use the GEKS procedure on all bilateral comparisons of the form leading to:

$$PPP_{jk}^{Jevons-GEKS(S)} = \prod_{i=1}^{C} \left[ PPP_{j'i}^{Jevons(S)} \cdot PPP_{i'k}^{Jevons(S)} \right]^{1/C} \quad (4.14)$$

The resulting indices are transitive and based on the binary indices that take into account representativity of the items priced in different countries.

The Eurostat-OECD method represents a viable approach that makes use of the representativity status of products priced in different countries. A few points of interest may be noted here. First, in a given binary comparison, the procedures described in Sections 4.5.2 and 4.5.3 make use of data corresponding to those items that are representative in one country which are also priced in the other country. In this process there is the possible loss of information. For example data on representative price items in country $j$ which are not priced in country $k$ do not enter the computation. Similarly prices of products which are considered unrepresentative products in both countries do not influence the binary comparison. Second, a more important one, is that the Jevons-GEKS* and Jevons-GEKS*(S) do rely heavily on price comparisons for commodities which are representative in one country but not in the other. Intuitively, such comparisons tend to be distorted as the commodity is representative in one country but not representative in the other. There is no guarantee that these distortions cancel each other\textsuperscript{22}. In that case, it is likely that these distortions may in fact be cumulating and the resulting BH parities may be highly distorted. This may not be a major problem when comparisons are made in a region where all the countries are fairly similar and there is a significant overlap of products. The Eurostat-OECD addresses this issue through the use of equi-representativity which endeavours to equalize the number of representative products in each country. third, it is difficult to

---

\textsuperscript{22} This is easily true in the hypothetical case where there are no items priced which are considered representative in both countries.
generalize the Jevons-GEKS(*) index to attach different weights to the representative and unrepresentative items in the BH. Finally, the Eurostat-OECD approach which uses the Jevons index requires one single price observation for each item from each country. This is not a major restriction as it is common practice to use the national annual average price of the item as input into the BH PPP computations. However, in the event where each country provides all the price quotations, the Eurostat-OECD approach cannot make direct use of the price quotations and these detailed data need to be aggregated into an average before it can be used. The regression based approach that underpins the country-product-dummy method solves this problem to a certain degree.

5. The Country-Product-Dummy (CPD) Method for the Computation of BH PPPs

The country-product-dummy (CPD) method was first introduced by Summers (1973) where he proposed a simple regression based method to fill missing price data in an incomplete tableau of prices at the BH level. The method was subsequently used in various phases of the ICP conducted by the research team at the University of Pennsylvania. Kravis, Heston and Summers (1982) report on Phase III of the ICP provides a detailed account of how the CPD method provided an aggregation method at the BH level. However, the OECD-Eurostat continued to use the GEKS methods and its variants through the last three decades. There has been renewed interest in the CPD method due to the recent work of Rao (1995, 2001, 2004, 2005, 2009), Diewert (2004, 2005) and also its recent use with PPP computations for poverty (Deaton, 2003).

This section describes the CPD method and show how the CPD method can be used in the computation of BH PPPs especially in the case where there is information regarding representativity status of items in different countries.

The Basic CPD Model

Following the notation used in Section 4.2, let $p_{ij}$ represent the price of item $i$ in country $j$ ($i=1,2,...,N; j=1,2,...,C$). It is useful to state the CPD model in the form that is directly relevant for international comparisons. The basic statistical model underlying the CPD method can be stated as:

$$p_{ij} = PPP_j P_i u_{ij} \quad \text{for} \quad j=1,2,...,C \quad \text{and} \quad i=1,2,...,N$$  \hspace{1cm} (4.15)

where $PPP_j$ is the purchasing power parity of currency of $j$-th country; $P_i$ is the international average price of $i$-th commodity and $u_{ij}$ are independently and identically distributed random variables. In the present chapter these disturbances are assumed to be lognormally distributed or that ln $u_{ij}$ s are normally distributed with mean 0 and a constant variance $\sigma^2$. Several features of the CPD model are noteworthy.

First, prices used in the CPD model may be considered as a single price observation for each item in each country where it is priced. The CPD model is general enough to accommodate the case where there are several price quotations available for each commodity in each country, a case considered in Diewert (2004). When individual price quotations are used it would be possible to extend the CPD model to incorporate additional characteristics associated with the quotation including information on the type of outlet and on the rural/urban location for the transaction.

\footnote{For example, it would be difficult to accord a weight of 3 to price observation for a representative or important product and a weight of 1 to unrepresentative products. This will require a further modification of the Sergeev (2003) suggestion.}
Second, in the ICP only single price observations representing the annual average prices of items in the BH are used. If there is information regarding the standard error associated with the average price, then this information can be incorporated into the model using different variances for different products.

Third, the CPD model in (4.15) is usually referred to as the law of one price reflected by a single average price for a commodity across all the countries and a single measure of price level for each country represented by PPP_j.

Finally, the CPD model can be best described as a hedonic regression model where the characteristics used are the country and the commodity specifications. In order to write it in a standard hedonic model, the CPD model can be written using logarithmic prices. Taking natural logs on both sides, the model can be written as:

\[
\ln p_{ij} = \ln PPP_j + \ln P_i + \ln u_{ij} = \alpha_j + \gamma_i + v_{ij}
\]

where \(v_{ij}\) are random disturbance terms which are independently and identically (normally) distributed with zero mean and variance \(\sigma^2\). The CPD model can be seen as a simple fixed effects model where country-effects provide estimates of purchasing power parities and commodity-specific effects provide estimates of international prices.

The parameter \(\alpha_j\) is interpreted as the general price level in country \(j\) relative to prices in other countries included in the comparison. It is possible to express \(\alpha_j\) relative to a reference country (say country 1), then \(\alpha_j\) represents the purchasing power parity of country \(j\) showing the number of country \(j\) currency units that have the same purchasing power as one unit of currency of country 1 or the reference country.

Then PPP for country \(j\) is given by:

\[
PPP_j = \exp\left(\hat{\alpha}_j\right)
\]

As the estimated PPP depends upon the estimated parameter values, it is possible to derive the standard errors associated with \(PPP_j\) which is not possible when the jevons method discussed in Section 4.4 is used.

**The CPD Regression Model**

The simple model in (4.16) derives its title the country-product-dummy (CPD) method as it can be expressed as a regression equation where all the explanatory/ regressor variables are essentially dummy variables (one for each country and one for each commodity). The basic model

\[
\ln p_{ij} = \alpha_j + \gamma_i + v_{ij}
\]

can be written as

\[
y_{ij} = \ln p_{ij} = \alpha_1 D_1 + \alpha_2 D_2 + \ldots + \alpha_c D_c + \eta_1 D_1^* + \eta_2 D_2^* + \ldots + \eta_N D_N^* + u_{ij}
\]

\(^24\) Model in (4.16) is not identified and it requires normalization before the parameters of the model can be estimated.
where \( D_j (j=1,2,\ldots,C) \) and \( D_i^* (i=1,2,\ldots,N) \) are respectively country and commodity dummy variables. Equation (4.18) can be written as

\[
y_{ij} = x_{ij} \beta + v_{ij}
\]

where \( x_{nc} = [D_1 \ D_2 \ D_C \ D_1^* \ D_2^* \ D_N^*] \) and \( \beta = (\alpha_1 \ \alpha_2 \ \alpha_C \ \gamma_1 \ \gamma_2 \ \gamma_N) \)

where the values of the dummy variables are determined at the \( ij \)-th observation.

The main advantage of the CPD model in (4.18) is that it is possible to make use of very sophisticated econometric tools to derive interesting results. See Rao (2004) for more details as to how the the CPD model can be used in dealing with a number of data related issues are discussed below.\(^{25}\)

**CPD Method with Complete Price Tableau and with item specific weights**

Now consider the case where all the items in the BH are priced in all the countries. This is the case that corresponds to the case where Jevons index is used when all the countries price the same set of products.

In this case, for the aggregation at the BH level where there no weights, parameters \( \alpha_j \) and \( \eta_i \) can be estimated using simple unweighted or ordinary least squares by minimizing

\[
\sum_{i=1}^{N} \sum_{j=1}^{C} (\ln p_{ij} - \alpha_j - \gamma_i)^2
\]

(4.19)

The first order conditions for optimization with respect to \( \alpha_j \) and \( \eta_i \) lead to the following system of \( C+N \) equations in as many unknowns.

\[
\alpha_j = \frac{1}{N} \sum_{i=1}^{N} \ln p_{nc} - \sum_{i=1}^{N} \gamma_i \quad \text{for} \quad j=1,2,\ldots,C \quad \text{and}
\]

\[
\gamma_i = \frac{1}{C} \sum_{j=1}^{C} \ln p_{nc} - \sum_{j=1}^{C} \alpha_c \quad \text{for} \quad n=1,2,\ldots,N
\]

This system can be solved by imposing a linear restriction on the unknown parameters. For example, if \( \alpha_1 = 0 \) is the restriction imposed, it can be easily shown that, for each \( j=2,\ldots,C \)

\[
\hat{\alpha}_j = \frac{1}{N} \sum_{i=1}^{N} \left[ \ln p_{nj} - \ln p_{n1} \right] \quad \text{or} \quad \text{PPP}_j = \exp(\hat{\alpha}_j) = \prod_{i=1}^{N} \left[ \frac{p_{ij}}{p_{i1}} \right]^\frac{1}{N}
\]

(4.20)

\(^{25}\) Hill and Syed (2010) demonstrate how the CPD model can be used along with individual price quotations from different countries to obtain estimates of rural-urban price differentials and the outlet effects.
Using the solution in (4.20), comparisons of price levels between two countries $j$ and $k$, represented by can $PPP_{jk}$ be derived as:

$$PPP_{jk} = \exp(\hat{\alpha}_j) \frac{N}{N} \prod_{i=1}^{N} \left[ \frac{p_{jk}}{p_{ij}} \right]^{1/N} \quad (4.21)$$

The $PPP_{jk}$ obtained using the CPD model given in (4.21) is identical to the Jevons index presented in equation (4.4) in Section 4.4. As in the case of Jevons index, the index in (4.21) is obviously transitive and base invariant. The only difference is that as the CPD method makes use of a regression model it is possible to derive the standard error associated with each $PPP_{jk}$. It was shown in Rao (2004) that the estimated variance of $PPP_{jk}$ is given by

$$Est\,Var(\hat{\alpha}_j) = \frac{2}{N} \hat{\sigma}^2$$

where $\hat{\sigma}^2$ is an unbiased estimator of $\sigma^2$ and it is given by

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{C} \sum_{i=1}^{N} e_{ij}^2}{CN - (C + N - 1)} \quad (4.22)$$

where $e_{ij} = \ln p_{ij} - \hat{\alpha}_j - \hat{\gamma}_i$ is the least squares residual. Using (4.22) the estimated variance of

$PPP_j$ with a numeraire country, say, country 1, is given by

$$Est\,Var(PP\hat{P}_j) \approx Est\,Var(\hat{\alpha}_j) \cdot (\hat{\alpha}_j)^2 \quad (4.23)$$

Equation (4.23) can be used in deriving estimated variance for PPPs with any other countries as the reference country.

**CPD with Incomplete Price Tableau**

The case where all the items are priced in all the countries is rarely encountered in practice. In fact the general experience in international comparison exercises is that only a few items are priced in each of the participating countries resulting in a rather sparse price tableau. This subsection examines the nature and role of the CPD method in this context and it is contrasted with the alternative aggregation method based on variants of the Jevons method used by the Eurostat-OECD.

The CPD model described in (4.16) and (4.18) can be used in conjunction with incomplete data provided that the price data set is connected as illustrated in Table 4.26. The CPD model and the least squares estimation shown in equations (4.19) and (4.20) go through with appropriate modifications. Rao (2004)

---

26 This is the case where not all the items in the BH are priced in all the countries.
provides algebraic expressions and the necessary proofs to support the following properties of the CPD method relative to the Jevons-GEKS approach described in Section 4.4.

1. The CPD and the GEKS methods provide identical estimates of PPPs when the price tableau is complete or equivalently when all the countries price the same set of products. There is no real problem of choice. However, the GEKS which expresses the $PPP_{jk}$ as a geometric mean of the price relatives for all the items provides no measure of reliability as in the case of the CPD approach.

2. When the price tableau is incomplete, the CPD and the Jevons-GEKS methods provide different numerical values. The CPD method makes use of all the price information in a single step whereas the Jevons-GEKS method uses the information in two stages. At the first stage binary comparisons are made using only prices of items that are priced in a given pair of countries. Obviously, data on prices of items which are priced in one country but not the other are ignored. An indirect use is made of the price data for the other items through the GEKS extension of the binary Jevons indices. Once again, no standard errors are available for the PPPs derived using the Jevons-GEKS method.

3. The point made above in can be further elaborated as follows. When the price tableau is incomplete, one can estimate the CPD model and fill in the missing price data. Then one has a complete tableau of prices. The CPD based PPPs remain unchanged if the CPD model is applied a second time after filling the missing prices. This is an indication that full use of the price data is made under the CPD method. However, the Jevons-GEKS applied to the incomplete tableau differ from the Jevons method applied to the tableau after the missing price data are imputed using the CPD model. This means that improvements can be made to Jevons-GEKS through the use of CPD fillers which, in turn, implies that the GEKS does not make full use of the price data in the incomplete tableau.

### CPD method with information on Representativity and Importance

In this subsection, two possible modifications of the simple CPD model discussed in (4.16) and (4.18) are discussed. The first case refers to the case where additional information on representativity of each item priced in each of the countries is available. In this case it is possible to extend the CPD regression model to directly account for the possible upward (or downward) bias caused by prices of unrepresentative items. The representativeness concept was used in the 2005 ICP. The second concerns the notion of importance attached to the items whereby each item is classified as important or unimportant in each of the countries. The notion of importance is adopted for use in the 2011 round of the ICP. Extensions of the CPD model in these two cases are discussed below.

### Country-Product-Representativity-Dummy (CPRD) Model

Recalling the discussion on representativity in Section 4.3 of this chapter as well as in Chapter 1, the basic idea is that representative products tend to be cheaper than unrepresentative products within a basic heading. This means that in addition to the country and product dimensions used in the CPD model there is an additional dimension of representativity considered to be critical and therefore included.

Accommodating representativity information is quite straightforward through the introduction of another dummy variable representing the additional dimension. In this case, for each price observation a Representativeness Dummy variable, $R$, is defined such that the value of $R = 0$ if the price observation

---

27 The opposite case where unrepresentative products are cheaper can be equally accommodated into the CPD model.
corresponds to a representative item and $= 1$ if the particular item is not representative. The basic CPD model in equation (4.18) may be extended to include the representativeness dummy as follows:\footnote{The model presented here is in a format slightly different to that used in Hill (2007) for the ICP handbook 2003-2006 (World Bank, 2006).}:

$$y_{ij} = \ln p_{ij} = \alpha_1 D_1 + \alpha_2 D_2 + \ldots + \alpha_c D_c + \eta_1 D_1^* + \eta_2 D_2^* + \ldots + \eta_N D_N^* + \delta R + u_{ij}$$

$$= \sum_{j=1}^C \alpha_j D_j + \sum_{i=1}^N \eta_i D_i^* + \delta R + u_{ij} \quad (4.24)$$

The parameters in the model can be estimated using the standard least squares after imposing the numeraire restriction setting one of the $\alpha_j$’s equal to unity. The resulting estimates of PPPs are essentially adjusted for the upward bias caused by those price observations that are not representative. It is expected that in the general case where unrepresentative are more expensive the estimate of $\delta$ will be positive:\footnote{In a recent study Hill and Syed (2010) present estimates of the representativity coefficients computed using data from a selected set of countries from the 2005 ICP Asia-Pacific and found that the coefficients could be negative for some basic headings and positive for the others and that no definite conclusions could be drawn. Their findings may partly reflect the difficulties the national price statisticians had in determining whether or not a particular product was representative in their countries.}

The CPRD model makes it much easier to handle the bias induced by prices of unrepresentative items through the magnitude of $\delta$. Such an adjustment is not possible in the case of the Jevons-GEKS procedure. In addition, it is clear that considerable price information is lost in the general architecture of the Jevons-GEKS* and Jevons-GEKS*(S) methods described in Section 4.4.

Based on the advantages of the CPRD model and as it makes use of all the information contained in the data set, the CPRD model was recommended for use in 2005 ICP for aggregation at the basic heading level. However, the CPRD model could not be used in the African and Asia-Pacific regions due to the problems associated with the determination of representative and unrepresentative products. The CPRD method was used in the Latin American comparisons. The Eurostat-OECD used the Jevons-GEKS*(S) method described in Section 4.4.

**CPD model with Importance weights attached to Price observations**

Given the practical problems associated with the identification of representativeness of items priced in different countries, the 2011 ICP has decided to use the notion of *importance* of the product priced. In particular, each item priced will be identified as either *important* or *unimportant*. The *importance* information will be used through the use of weights attached to price observations. In its meeting in April 2011, the Technical Advisory Group (TAG) of the ICP recommended that a weight of 3 be attached to products identified as important and a weight of 1 be attached to items that are deemed to be unimportant:\footnote{It was noted that unweighted use of price observations amounts to giving equal weight to products to products that are important as well as those that are unimportant.}

It is fairly simple to attach weights in the estimation of the parameters of the CPD model. This is equivalent to running weighted least squares in the place of simple unweighted least squares. Suppose $w_{ij}$ is the weight attached to the price quotation for $i$-th commodity in $j$-th country then the weighted least squares simply minimises
\[
\sum_{i=1}^{N} \sum_{j=1}^{C} w_{ij} \left( \ln p_{ij} - \alpha_{i} - \gamma_{i} \right) ^{2} = \sum_{i=1}^{N} \sum_{j=1}^{C} \left[ w_{ij} \left( \ln p_{ij} - \alpha_{1} D_{1} - \alpha_{2} D_{2} - \ldots - \alpha_{C} D_{C} - \eta_{1} D'_{1} - \eta_{2} D'_{2} - \ldots - \eta_{N} D'_{N} \right) \right]^{2}
\]

(4.25)

with respect to the unknown parameters which in turn result in estimate PPPs. The TAG recommendation is to use \( w_{ij} = 3 \) if the commodity is important or representative and \( w_{ij} = 1 \) if it is unimportant.

As an extension of this procedure, it may be possible to include information on approximate expenditure or sales shares in the place of arbitrarily stated weights. The main feature of the model in (4.25) is that it is exactly the model one would use if expenditure share weights were available. Rao (2009) discusses the expenditure share weighted CPD model.

The extensions and variations of the CPD model discussed in this section are limited to the cases applicable to the estimation of basic heading PPPs. It is clear that the CPD method offers major advantages over the alternative based on the Jevons index and variations of the Jevons-GEKS index. Given the existence of noise in the price data as the data are collected through price surveys, methods like the CPD based on statistical models are best suited to account for noise and also to provide estimates of reliability in the form of standard errors associated with the PPPs at the BH level.

**A Numerical Example**

A numerical example designed to illustrate various methods of aggregation used in deriving PPPs at the basic heading level is presented here. In particular, the properties of the Jevons-GEKS, Jevons-GEKS*, Jevons-GEKS*(S) and the CPD, CPRD and the weighted CPD methods are illustrated using a simple example. 31

The example refers to the Basic Heading Materials for Maintenance and Repair of the Dwelling in the ICP. There are four items listed under this BH and there are four countries in this illustration. In this example, all the items are priced in all the countries which is useful in illustrating the equivalence of the CPD and Jevons-GEKS methods in special cases. Representativity of different products are indicated through an asterisk (*) attached to the price quotation.

---

31 The author is greatly indebted to Sergey Sergeev for providing the data and the computations required for the numerical example.
Table 2: Price Data for Aggregation at BH Level

BH: Materials for Maintenance and Repair of the Dwellings

<table>
<thead>
<tr>
<th>Description</th>
<th>Qty</th>
<th>unit</th>
<th>Country 1</th>
<th>Rep</th>
<th>Country 2</th>
<th>Rep</th>
<th>Country 3</th>
<th>Rep</th>
<th>Country 4</th>
<th>Re</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paint, indoor use</td>
<td>10</td>
<td>l</td>
<td>33.88</td>
<td></td>
<td>34.90</td>
<td>*</td>
<td>753.36</td>
<td>*</td>
<td>89.45</td>
<td>*</td>
</tr>
<tr>
<td>Paint, outdoor use</td>
<td>10</td>
<td>l</td>
<td>49.19</td>
<td></td>
<td>71.34</td>
<td></td>
<td>1317.93</td>
<td>*</td>
<td>149.05</td>
<td></td>
</tr>
<tr>
<td>Silicone</td>
<td>300</td>
<td>g</td>
<td>4.54</td>
<td>*</td>
<td>5.29</td>
<td>*</td>
<td>84.74</td>
<td></td>
<td>7.54</td>
<td></td>
</tr>
<tr>
<td>Cement</td>
<td>25</td>
<td>kg</td>
<td>4.57</td>
<td></td>
<td>6.30</td>
<td>*</td>
<td>60.07</td>
<td>*</td>
<td>5.55</td>
<td></td>
</tr>
</tbody>
</table>

In this example all the four commodities are priced in all the countries. Therefore, the GEKS parities should be identical to those derived using the CPD model. In addition, only one item is considered representative in country 1 where as countries 2 and 3 have three products that are representative. In contrast, only one item is representative in country 4.

The following aggregation methods are considered in the following example:

1. Simple unweighted CPD model;
2. CPRD model with a representativeness dummy;
3. CPD model with weights for representative items equal to 3 and 1 for unrepresentative items similar to the TAG recommendation;
4. Jevons-GEKS* method where the representativeness is taken into account; and finally
5. Jevons-GEKS*(S) which gives additional weight to commodities which are representative in both countries.

The computed PPPs using different methods are presented below:

Table 3: Multilateral PPPs for BH by different methods

Materials for the maintenance and repair of the dwellings (4 items)

<table>
<thead>
<tr>
<th>Method of the calculation</th>
<th>Purchasing Power Parities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Country 1</td>
</tr>
<tr>
<td>Jevons-GEKS*</td>
<td>1.000</td>
</tr>
<tr>
<td>Jevons-GEKS without *</td>
<td>1.000</td>
</tr>
<tr>
<td>Jevons-GEKS*(S )</td>
<td>1.000</td>
</tr>
<tr>
<td>CPRD</td>
<td>1.000</td>
</tr>
<tr>
<td>CPD unweighted (weights 1:1)</td>
<td>1.000</td>
</tr>
<tr>
<td>CPD weighted (weights 3:1)</td>
<td>1.000</td>
</tr>
<tr>
<td>Exchange Rate (NC/Euro)</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Several features of the Table 4.3 are worth noting. First, PPPs for the EKS without using (*) is identical to the Unweighted CPD. This result is consistent with the analytical result which shows the equivalence of these two methods when the price tableau is complete. However, when representativity information is different methods lead to different results. Otherwise, there are no obvious trends. The weighted CPD method with 3:1 weighting scheme seems to perform quite well. The CPRD method appears to produce the lowest PPP values followed by the GEKS* method.

Given that the CPD method makes it possible to employ a range of econometric techniques on the price data used and in addition produces standard errors for the estimated PPPs, the use of weighted CPD appears to be a particularly good choice.

6. Linking Regions at the Basic Heading level in 2005 ICP and Looking Forward to 2011 ICP

The main objective of this section to outline the methodology used in linking PPPs at the BH level in the 2005 ICP and provide a brief overview of the methodology being considered in the 2011 ICP. The problem of linking PPPs above the basic heading level is considered in depth in Chapter 8 by Erwin Diewert. In essence, Chapter 8 makes use of the linked PPPs at the BH level as inputs into the aggregation process.

Linking BH Level PPPs in 2005 ICP

The 2005 ICP embraced a fully regionalised approach to the compilation of PPPs and international comparisons of real gross domestic product and its components. As a part of the regionalised approach, all the participating countries were classified into six regions, viz., Africa, Asia-Pacific, CIS, Eurostat-OECD, South America and West Asia. In the 2005 ICP, Egypt and Russia belonged to two different regions at the same time. Egypt participated in the African and West Asian regional comparisons whereas Russia participated in the Eurostat-OECD and the CIS regions. Two major steps were involved in the compilation of BH level PPPs for the 146 participating countries in the global comparison. First, price comparisons were undertaken in each of the regions coordinated by a regional coordinating agency working under the guidance of the ICP Global Office. At the conclusion of the project at the regional level, in each of the regions PPPs at the BH level were compiled for the 155 basic headings in the GDP comparisons. In each region, one of the countries was selected as the numeraire or reference country so that PPPs are expressed in the currency units of the reference country. In order to make use of these regional sets of PPPs within a global comparison it is necessary to express the regional BH PPPs in the currency units of a global numeraire. A major requirement in this process is the fixity requirement which stipulates that the relativities between purchasing powers of currencies of countries within a region remain unaltered in the process of conversion to a global numeraire. For a more complete description of the general approach used in making global comparisons, see Section 1.5 of Chapter 1 by Rao.

The methodology used in 2005 ICP to convert regional BH level PPPs maintaining fixity is described below. The essential steps involved in the process are:

- Compile PPPs for currencies of countries within each region using a regional numeraire currency for all the 155 BHs in the GDP
- A set of 18 ring countries were identified, with six countries from Africa, four from the Asia-Pacific and two each from the Eurostat-OECD and West Asian regions. The CIS region was linked using Russia as the bridge country.

See Chapter 2 of the ICP book for more details about the organisational structure of the ICP.
• All the ring countries conducted additional surveys to collect prices of items in the ring product list. The ring list was developed by the Global Office after examining the regional product lists. This process was used only for Household Consumption. For all the other aggregates, regional comparisons were based on a global list of items.

• The price data collected by the 18 ring countries were used in compiling linking factors used to convert the regional numeraire currency units into the global numeraire currency. The US dollar was used as the global numeraire.

• The methodology used for linking both at the BH level and at higher levels of aggregation were developed by Diewert and are described in detail in Diewert (2008, 2010b).

The aggregation methodology used in the computation of linking factors is described below.

Notation
Suppose that there are \( R \) regions in the comparison with \( C(r) \) for \( r = 1, 2, \ldots, R \) countries in region \( r \). In the 2005 ICP, \( R = 6 \) and Table 1.3 of Chapter 1 by Rao shows the distribution of countries by the region. Let \( L(r) \) \((r=1,2,\ldots,R)\) be the number of link countries a total of \( L = \sum_{r=1}^{R} L(r) \). In 2005 \( L = 18 \).

The Method
Let \( PPP_{rcn} \) represent the PPP for the \( n \)-th basic heading in country \( c \) belonging to region \( r \). Without loss of generality, let the first country in each region be the numeraire country. The following steps are used in the computation of linking factor for the \( n \)-th basic heading.

Step 1: Consider all the ring countries in region \( r \). Let \( p_{ircn} \) represent the price of the \( i \)-th item priced in link country \( c \) in region \( r \). These are expressed in the national currency units of country \( c \).

Step 2: Convert all the prices of ring list items in \( n \)-th basic heading in link country \( c \) of the in region \( r \) into numeraire or reference country currency units using:

\[
p_{ircn}^{*} = \frac{p_{ircn}}{PPP_{rcn}} \tag{4.26}
\]

This means that the price in the link country is converted using \( PPP_{rcn} \). For example, if PPP for the Rice basic heading in Sri Lanka in the Asia-Pacific region is 5.85 to the HK dollar which was the numeraire currency in the Asia-Pacific region, then prices of ring products belonging to the Rice basic Heading collected from Sri Lanka are all divided by 5.85 thus expressing ring product prices in HK dollars. At the end of this step, prices of all the ring list items priced by ring countries in the Asia-Pacific region are all expressed in HK dollars.
Step 3: Use the converted ring list prices for all the ring countries in a CPD regression model to derive PPPs for the numeraire currencies in different regions. The CPD model used can be written as:

\[
\ln p_{rcn}^L = \alpha_{rn} + \beta_n + \mu_{rcn} \tag{4.27}
\]

where \(\exp(\alpha_{rn})\) represents the purchasing power parity of the currency of region \(r\) in terms of currency units of the global numeraire or reference currency (say the US dollar) for the \(n\)-th basic heading.

Step 4: Apply the CPD regression in (4.27) in Step 3 for all the 155 basic headings.

Step 5: Convert the BH PPPs within each region expressed in the reference currency of the region (eg. HK dollar in the Asia Pacific) into the global numeraire using the PPPs derived in Step 4. At the end of this step, the linked PPPs at the BH level are given by:

\[
PPP_{rcn}^L = PPP_{rcn} \cdot \exp(\alpha_{rcn}) \tag{4.28}
\]

For example, if the BH PPP for Rice in Sri Lanka in the Asia Pacific comparison is 5.85 Sri Lankan rupees for one HK dollar and if the linking factor computed using ring prices leads to say 6 HK dollars per one US dollar, then the PPP for the Rice basic heading for Sri Lanka in the global reference currency, US dollar, is 35.10 Sri Lankan rupees per US dollar.

A numerical example of the computation of the linking factors using illustrative data can be found in Appendix H of the Global ICP Report, World Bank (2008).

Sergeev (2009) raises an interesting issue that arises due to the presence of different number of ring/link countries from different regions, i.e., \(L(r)\) varies with \(r\). In ICP 2005, there were 6 countries from Africa and 4 countries from Asia compared to 4 from the OECD\(^{33}\). The main point is that within the CPD framework, it can be shown that the international average prices are averages of prices from the linking countries and therefore regions with greater number of linking countries may be seen to exert a larger influence.\(^{34}\) However, the rationale for including more than one country of a region in the ring list is that for those regions that are large and those that exhibit a large variation in prices, it is necessary to use more price data drawn from a diverse set of countries representing the region. There are two possible solutions to the question raised in Sergeev (2009).

(i) Express the uncertainty attached to prices from a region through a larger variance for the corresponding disturbance term in the CPD model and then apply generalised least squares method which accords less weight to those observations with larger variance. This eliminates the problem discussed by Sergeev. However, it is difficult to know the extent of variability in prices across countries within a region.

(ii) Alternatively, the suggestion made in Sergeev’s paper may be implemented. Instead of using country specific prices for each of the link countries in a region, simply take the geometric mean of the prices after converting them into the currency of the numeraire country. In this case, we simply use

\(^{33}\) The full list of ring countries for the ICP 2005 round consists of: Brazil; Cameroon; Chile; Egypt; Estonia; Hong Kong; Japan; Jordan; Kenya; Malaysia; Oman; Philippines; Senegal; Slovenia; South Africa; Sri Lanka; United Kingdom; and Zambia.

\(^{34}\) Equations (4.26) to (4.28) in Sergeev (2009) can be derived from the normal equations associated with the ordinary least squares method.
In the next step, use these geometric means as inputs into the CPD model to generate regional linking factors described in equation 4.27. Use of the geometric mean ensures that each region is represented by a single vector of prices in the CPD regression model thus ensuring that no region exerts more than proportionate influence on the basic heading level parities that are used in deriving linking factors at higher levels of aggregation.

Given that different regions exhibit different levels of variability and the use of geometric average prices from ring countries entails a loss of information on prices, it is appropriate to continue with the current practice of use of different number of ring countries in the BH computations.

An additional comment on the possible use of representativity and importance information in the derivation of the linking factors may be made here. As the CPD regression in equation (4.27) makes use of price observations from a number of countries within each region and given that each country may have a different representativity status for a product in the ring list, it would be possible to make use of such information in deriving the regional linking factors for different basic headings. However, use of representativity information would not be possible if Sergeev (2009)’s suggestion to use the regional average price in (4.29) is implemented. It is difficult to attach any meaning for representativity for a price that is the average over a number of different countries.

**A New Approach based on Core Products in 2011 ICP**

An evaluation of the ring country approach used in the 2005 ICP has raised a number of issues. These concern the choice of ring countries from each of the regions and the likely influence of a less than ideal choice of reference countries on the linking factors that have a direct impact on the global PPPs at the GDP level and the resulting real income comparisons. The construction of the ring product lists and pricing of these products in a limited set of ring countries from each region were also considered to be less than satisfactory.

As an improvement over the 2005 methodology it was decided that the 2011 regional comparisons will be linked through price data collected for set of products identified as core products. The core product list is supposed to represent the products used in both the developed and developing countries. A list in excess of 600 core products was prepared by the Global Office. The core products are expected to strengthen the link between regional and global comparisons through the following steps:

- The core products are included in the product lists of all the regions;
- Regions are encouraged to price as many core products as possible in their regional price surveys;
- Regions will make use of price data on core products as well as the region-specific product lists in the computation of PPPs for different BHs;
- Linking factors to link regional BH PPPs will make use of prices of core products collected by all the participating countries in all the regions and not on prices collected for a selected set of ring countries; and the
CPD regression in (4.27) is based on prices of all the core products in all the countries participating in the 2011 round of the ICP. The number of countries is expected to be around 180. Use of price data from all the countries is likely to produce more robust and reliable set of linking factors that are likely to improve the quality of the global comparisons.

Exploratory empirical analyses using core product prices collected in the first two quarters of the 2011 ICP will be conducted to examine the nature and reliability of the linking factors resulting from the new approach. In the meanwhile, research is continuing into alternative ways of using core product prices. Hill (2011) proposes methods to use the core product prices in computing PPPs at and above BH levels imposing within-region fixity in the ICP.

7. Conclusions

The Jevons-GEKS and the CPD are the two main aggregation methods for the compilation of PPPs at the basic heading level. When all the items are priced in all the countries and if all the items are considered equally important, then these two methods lead to the same numerical values for the estimated PPPs. However, the CPD method has the advantage of providing standard errors which can be used as measures of reliability. In the case where some products are representative or important, then three variants of the GEKS method and two variants of the CPD method are available. On a conceptual level, the weighted CPD with weights reflecting the importance of the product is superior to the use of the CPRD which relies on the existence of a systematic bias induced by unrepresentative products. In terms of the use of the Jevons-GEKS and its variants, in general the use of Jevons-GEKS based methods appear to discard some price data which is not the case when the CPD based methods are used. The Jevons-GEKS*(S) due to Sergeev (2003) appears to perform well among the GEKS based methods.

Looking forward to the 2011 ICP, the Jevons-GEKS and the CPD based methods need to be reexamined in the light of the decision to use a list of “core products” that are going to be priced in all the countries in all the regions. The use of core products is designed to eliminate the reliance on a few selected ring countries for the purpose of linking regions. A number of methods are currently being developed and being discussed at various meetings of the Technical Advisory Group. The CPD method is ideally placed for the current approach of using prices of the core product prices collected from all the countries in the global comparisons. The Jevons-GEKS method would be of limited applicability in this case as the method can be used only if average prices representing each of the regions are available. However, use of average prices implies a loss of information. While search is continuing to identify suitable aggregation methods to link regional BH PPPs using the core product price data, in the interim the CPD method is currently best suited to make use of all the core prices collected from all the countries.
References


