

Improving the ICP inter-regional linking procedure

February 27,

2025

Yuri Dikhanov (World Bank)

In this paper, we look at the problems associated with calculating interregional Purchasing Power Parity (PPP) indexes at the Basic Heading level. We focus on variants of the Country Product Dummy (CPD) method for linking countries that was originally proposed by Summers (1973), and then extended by Diewert (2008) to linking regions.

1. Algebra for the Country Product Dummy Multilateral PPPs

First, we present the algebra for the CPD method. We must mention that the CPD method was developed by Summers (1973) to deal with price matrices with missing data and for gap-filling purposes. The underlying assumption was that the prices in different countries should be approximately proportional. We follow the exposition in Diewert (2022a; 54-56), adapting his time series presentation of the method to the problem of comparing price levels across countries instead of across time periods.¹

To start, consider a simple case with only one region. Suppose that there are K countries in the international comparison of prices with N products in scope. If product n is priced in country k for the time period under consideration, denote its average price by p_{kn} for $k = 1, \dots, K$ and $n \in S(k)$ where $S(k)$ denotes the set of products that are priced in country k . The price is expressed in units of country k 's currency. The basic assumption made in the CPD model is that the observed country prices satisfy the following equations (approximately):

$$(1) \quad p_{kn} \approx \pi_k \alpha_n ; \quad k = 1, \dots, K; n \in S(k);$$

where α_n is the world price² for product n and π_k is the overall level of prices (for the N product prices in scope) in country k relative to the level of prices in other countries. Thus the basic assumption is that product prices vary proportionally across countries for the group of products in scope³. The parameter π_k is country k 's Purchasing Power Parity (PPP).

Take logarithms of both sides of equations (1) and add error terms to obtain the following linear regression model:

$$(2) \quad \ln p_{kn} = \rho_k + \beta_n + \varepsilon_{kn} ; \quad k = 1, \dots, K; n \in S(k);$$

where ε_{kn} is an error term and ρ_k and β_n are the logarithms of π_k and α_n ; i.e.,

$$(3) \quad \rho_k \equiv \ln \pi_k \text{ for } k = 1, \dots, K \text{ and } \beta_n \equiv \ln \alpha_n \text{ for } n = 1, \dots, N.$$

¹ His paper has more details and references to the literature.

² Diewert (2022a) uses the term “quality adjusted parameter” in the CPI context. In the ICP setting, the world product price seems more appropriate, as the CPD procedure allows for generating distinct average product prices of, say, coffee and tea in Coffee and Tea Basic Heading, or prices of coats and socks in Garments, Mercedes cars vs. Toyota vehicles, etc. As we will see later, the world product prices are expressed in the base country's currency.

³ This assumption is consistent with the requirement of the ICP that each basic heading includes more or less homogeneous products.

Estimates for the parameters ρ_k and β_n in equations (2) can be found by solving the following least squares minimization problem:

$$(4) \min_{\rho, \beta} \sum_{k=1}^K \sum_{n \in S(k)} [\ln p_{kn} - \rho_k - \beta_n]^2 = \min_{\rho, \beta} \sum_{n=1}^N \sum_{k \in S^*(n)} [\ln p_{kn} - \rho_k - \beta_n]^2$$

where $\rho \equiv [\rho_1, \dots, \rho_K]$ and $\beta \equiv [\beta_1, \dots, \beta_N]$ and $S^*(n)$ is the set of countries k that have priced product n for $n = 1, \dots, N$. Note that there are two equivalent ways of writing the least squares minimization problem. Solutions ρ and β to problem (4) will satisfy the following first order conditions for solving the minimization problem:

$$(5) \sum_{n \in S(k)} \ln p_{kn} = \sum_{n \in S(k)} [\rho_k + \beta_n] ; \quad k = 1, \dots, K;$$

$$(6) \sum_{k \in S^*(n)} \ln p_{kn} = \sum_{k \in S^*(n)} [\rho_k + \beta_n] ; \quad n = 1, \dots, N.$$

Let $N(k)$ equal the number of products that are priced in country k for $k = 1, \dots, K$ and let $N^*(n)$ equal the number of countries that price product n for $n = 1, \dots, N$. Using these definitions plus definitions (3) (to switch from the ρ_k and β_n to the π_k and α_n), equations (5) and (6) can be rewritten as follows:⁴

$$(7) \pi_k = \prod_{n \in S(k)} [p_{kn}/\alpha_n]^{1/N(k)} ; \quad k = 1, \dots, K;$$

$$(8) \alpha_n = \prod_{k \in S^*(n)} [p_{kn}/\pi_k]^{1/N^*(n)} ; \quad n = 1, \dots, N.$$

A solution $\pi \equiv [\pi_1, \dots, \pi_K]$ and $\alpha \equiv [\alpha_1, \dots, \alpha_N]$ to equations (7) and (8) will not be unique since $\lambda\pi$ and $\lambda^{-1}\alpha$ will also be a solution for any positive scalar λ . Thus, we are allowed to make one normalization on the π_k and α_n in order to obtain a unique solution. If we choose the normalization $\pi_1 = 1$, then we are choosing country 1 as the world numeraire country. From equations (8), it can be seen that α_n can be interpreted as a PPP adjusted world price for product n with the same normalization as the one on the π_k .

It is clear that π_k / π_m (the ratio of the PPP for country k to the PPP for country m) can still be estimated even when there aren't any matched product prices between the two countries. In this case the PPP for those countries will rely solely on indirect links⁵.

An important property of equations (7) and (8) is that they are equivalent to an unweighted geometric version of the Geary-Khamis system. In log terms the system becomes:

$$(7a) \rho_k = \sum_{n \in S(k)} [\ln p_{kn} - \beta_n] / N(k) ; \quad k = 1, \dots, K;$$

$$(8a) \beta_n = \sum_{k \in S^*(n)} [\ln p_{kn} - \rho_k] / N^*(n) ; \quad n = 1, \dots, N.$$

However, in the ICP, starting from 2011 a set of weights was implemented to reflect importance of individual products ("importance weights"). The corresponding modified procedure is called the CPD weighted (CPDW), and is the solution to the minimization problem (4) with weights w_{kn} :

$$(4a) \min_{\rho, \beta} \sum_{k=1}^K \sum_{n \in S(k)} w_{kn} [\ln p_{kn} - \rho_k - \beta_n]^2 = \min_{\rho, \beta} \sum_{n=1}^N \sum_{k \in S^*(n)} w_{kn} [\ln p_{kn} - \rho_k - \beta_n]^2$$

The GK-like equivalent to system (7a-8a) thus becomes:

⁴ The sets $S(k)$ and $S^*(n)$ are assumed to be nonempty for $k = 1, \dots, K$ and $n = 1, \dots, N$.

⁵ In the time-series this case would be equivalent to a complete churn of products which can happen in Electronics or Computers.

$$\begin{aligned}
 (7b) \quad \rho_k &= \sum_{n=1}^N w_{kn} [\ln p_{kn} - \beta_n] / \sum_{n=1}^N w_{kn} ; & k &= 1, \dots, K; \\
 (8b) \quad \beta_n &= \sum_{k=1}^K w_{kn} [\ln p_{kn} - \rho_k] / \sum_{k=1}^K w_{kn} ; & n &= 1, \dots, N.
 \end{aligned}$$

If a particular price is missing, the corresponding weight will be zero and that missing price will be excluded from computation⁶. The CPDW procedure was used in 2011 and 2017 ICP for the regional PPP computations with importance weights {3,1}: more “important” products were given weight 3, and less “important” – weight 1. It is understood that it would be impossible to collect real weights for individual products, and even if it were possible, it would be a meaningless exercise because the products collected in the ICP are representative products (varieties) of a huge universe of product varieties available across the globe. For example, while only about a dozen models of personal vehicles with particular specifications are priced, there exist many thousands of them. Similar situation is with consumer electronics and many other product categories. Much worse is the situation with garments. One solution would be to break down Basic Headings even further but that would not be feasible in practical terms. Thus, the importance weight represents an expert opinion on the relative importance of the item to determine price level within that Basic Heading.

However, the system (7b-8b) can incorporate any weights, including those above the Basic Heading level (see Annex I).⁷

In this case the whole PPP calculation process would use the same method throughout, starting from aggregating product prices into Basic Heading PPPs and then aggregating them up to the GDP level.

2. Extension of the CPDW to inter-regional linking

In ICP 2005, a new concept was introduced – the Ring countries (see Diewert (2008) “New Methodology for Linking the Regions” for a detailed exposition).

A group of ring countries (18 countries from five ICP regions) collected prices from a common list and this price information was used to link the regional basic heading prices across the 6 regions. However, since the CIS region was locked into the OECD/Eurostat region, in practice, there were only 5 regions to link, with the CIS, OECD and Eurostat countries forming a single region⁸.

The 2005 inter-regional linking procedure was an adaptation of the unweighted CPD model. Two additional steps were introduced in the model: in the first step a set of intra-regional PPPs was calculated using the method adopted in that region⁹. In the second step product prices of that region get divided by those intra-regional PPPs to convert all product prices into a regional numeraire.

Thus, if the within region parities are known, then prices in each region can be divided by the appropriate regional parity for that country in that region, and these *regionally adjusted prices* can be used as inputs

⁶ In fact, any price with weight zero will be excluded from computation.

⁷ Rao (1995) [*“On the Equivalence of the Generalized Country-Product Dummy (CPD) Method and the Rao System for Multilateral Comparisons”*, Working Paper No. 5, Centre for International Comparisons, University of Pennsylvania] described such a variation of that system (although not the same system as the systems discussed in this paper) for the above Basic Heading aggregation. A system similar to Rao’s was also proposed by Diewert.

⁸ A similar linking arrangement was used in 2011 and 2017 as well.

⁹ All regions use the CPD method [or its variant the CPRD with $p_{kn} \approx \pi_k \alpha_n \delta_u$ and an additional binary representativity variable δ_u , due to Cuthbert (1988), *“On Aggregation Methods of Purchasing Power Parities”*, Working Paper No. 56, November, Paris: OECD], except for OECD/Eurostat and CIS that use the GEKS-Jevons procedure (with representativity).

into the usual CPD model that has now only the regional log parities and the commodity adjustment factors (world product prices) as unknown parameters to be estimated.

We will modify system (7a-8a) to accommodate the regions.

First, we introduce some additional definitions:

Let $S^*(r,n)$ denote the set of countries k that have priced product n in region r for $n = 1, \dots, N$ and $r = 1, \dots, R$; and $S(r,k)$ denote the set of products that are priced in country k in region r for $k = 1, \dots, K(r)$ and $r = 1, \dots, R$;

Further, recasting the definitions from Chapter 1:

$\ln p_{rkn}$ is log of price of product n in region r , country k , adjusted by intra-regional PPP;

ρ_r is log (π_r) , or log of PPP for region r ($r = 1, \dots, R$);

β_n is log (α_n) , or log of world price for product n ;

R is number of regions;

$K(r)$ is number of countries in region r .

Now we can rewrite system (7a-8a) when applied to regions as follows:

$$\begin{aligned} (7c) \quad \rho_r &= \sum_{k=1}^{K(r)} \sum_{n \in S(r,k)} [\ln p_{rkn} - \beta_n] / N(r); & k = 1, \dots, K(r); \quad r = 1, \dots, R; \\ (8c) \quad \beta_n &= \sum_{r=1}^R \sum_{k \in S^*(r,n)} [\ln p_{rkn} - \rho_r] / N^*(n); & n = 1, \dots, N. \end{aligned}$$

where

$N(r,k) \equiv$ the number of products priced in country k of region r ; $r = 1, \dots, R$; $k = 1, \dots, K(r)$;

$N(r) \equiv \sum_{k=1}^{K(r)} N(r,k)$ is the total number of products priced in region r ; $r = 1, \dots, R$;

$N^*(r,n) \equiv$ the number of countries in region r that priced product n ; $r = 1, \dots, R$; $n = 1, \dots, N$;

$N^*(n) \equiv \sum_{r=1}^R N^*(r,n)$ is the total number of countries that priced product n ; $n = 1, \dots, N$.

System (7c-8c) was used in ICP 2005 for inter-regional linking. Having learned from the experience of ICP 2005 round, the approach described by system (7c-8c) was modified in two important ways for ICP 2011: (1) now all countries could price the global product list, i.e., they essentially became what the Ring countries were in 2005, and (2) importance weights $\{3,1\}$ were introduced.

Using the GK-style representation¹⁰ system (7c-8c) modified for 2011 then becomes:

$$\begin{aligned} (9) \quad \rho_r &= \sum_{k=1}^{K(r)} \sum_{n=1}^N w_{rkn} [\ln p_{rkn} - \beta_n] / \sum_{k=1}^{K(r)} \sum_{n=1}^N w_{rkn}; & k = 1, \dots, K(r); \\ (10) \quad \beta_n &= \sum_{r=1}^R \sum_{k=1}^{K(r)} w_{rkn} [\ln p_{rkn} - \rho_r] / \sum_{r=1}^R \sum_{k=1}^{K(r)} w_{rkn}; & r = 1, \dots, R; \quad n = 1, \dots, N. \end{aligned}$$

where w_{rkn} is weight of product n in region r , country k .

Because $\sum_{k=1}^{K(r)} \sum_{n=1}^N w_{rkn} [\ln p_{rkn} - \beta_n] / \sum_{k=1}^{K(r)} \sum_{n=1}^N w_{rkn} = \sum_{n=1}^N \sum_{k=1}^{K(r)} w_{rkn} [\ln p_{rkn} - \beta_n] / \sum_{n=1}^N \sum_{k=1}^{K(r)} w_{rkn}$,

¹⁰ To prevent spurious results from a GK system, a check for isolates (unlinked countries) needs to be performed:

s1. start with country $c(1)$ and item $p(1)$, check if any other country prices $p(1)$, if not then go to next item until match is found, if no such country is found – FAIL, if found – add $c(1)$ & $c(n2)$ to set S of linked countries, go to next step;

s2. for country $c(n2)$, check if a matched country, except those already in S , can be found for any item, if not – FAIL, if found – add $c(n3)$ to S ;

s3. repeat step s2 until no more matches can be found, if S is incomplete – FAIL; if S is complete – PASS.

If FAIL is recorded, - GK has isolate(s) and PPPs cannot be estimated for the system.

system (9-10) can be rewritten as:

$$\begin{aligned} (11) \quad \rho_r &= \sum_{n=1}^N w_n(r) [\ln p_n(r) - \beta_n] / \sum_{n=1}^N w_n(r); & r = 1, \dots, R; \\ (12) \quad \beta_n &= \sum_{r=1}^R w_n(r) [\ln p_n(r) - \rho_r] / \sum_{r=1}^R w_n(r); & n = 1, \dots, N. \end{aligned}$$

where

$$\begin{aligned} (13) \quad \ln p_n(r) &= \sum_{k=1}^{K(r)} w_{rkn} \ln p_{rkn} / \sum_{k=1}^{K(r)} w_{rkn}, \text{ or log of weighted geomean price of product } n \text{ in region } r; \\ (14) \quad w_n(r) &= \sum_{k=1}^{K(r)} w_{rkn}, \text{ or total weight of product } n \text{ in region } r. \end{aligned}$$

System (11-12) is much simpler than system (9-10) and is in effect essentially system (7b-8b) where countries are replaced by regions. Expression (13) describes the weighted geomean of prices of individual countries within a region (adjusted by the regional PPP), whereas expression (14) refers to total weight of individual products in each region. The unweighted case is described with similar formulae (see Annex II).

Thus, in our analyses we can concentrate on regional prices and regional weights, treating regions as entities. This makes analysis much simpler as we need to look at price matrices with dimensions [5; 631] instead of [146; 631].

Nature of the regional price

It is important to investigate what exactly is the regional price described by formula (13). The regional price is *not* an average of prices of individual countries per se. The regional price is in effect the price of the base country for that region, adjusted for inconsistencies of the base country's position vis-à-vis other countries in the region using the core list items only, versus the base country's position in the regional comparison, using all items used to generate the regional PPPs. The items used in the regional calculation may or may not include the core list items.

Consider *Rice* in Asia: the base country – Hong Kong – priced only two products from the global core list, and both products had relatively high prices within Asian regional comparison: the CPD residuals for those two items were 60% and 45% in log terms. Thus, if not adjusted for the regional PPPs (or if the regional item set would coincide with the global core list), those two products would significantly raise the regional PPP for Asia. However, the linking procedure is quite robust versus price outliers. For example, if we remove those two prices from the global linking altogether, the *Rice* linking factor (PPP) for Asia would change insignificantly: from 5.19 to 5.15¹¹.

The regional PPPs described by expression (11) can be understood as the linking factors that link the regional results expressed in regional currency (Hong Kong Dollars in case of Asia) to the global results expressed in the world numeraire currency (the US Dollars). Thus, the PPP of Bangladesh in the global comparison will be equal to the PPP of Bangladesh in the regional comparison multiplied by the Asian regional PPP (linking factor).

On average, the inconsistencies between regional and global core lists are quite insignificant: they are within 1% for four regions, and for Asia it is 3%. Some Basic Headings do exhibit significant variations, for example, *Rice* in Asia. However, even if using only the base country GCL prices in linking (i.e., linking only via the five regional base countries, and disregarding information from all other countries, thereby reducing the linking to the minimal Ring of five countries, so to speak), the resulting regional linking factors

¹¹ An important property of the linking procedure is that the regional base countries are not required to have any products priced from the global core list.

(PPPs) would not change significantly at the aggregate level, even though they may change substantially for individual Basic Headings. This is a positive result as it reinforces the outcome that global core list prices are highly correlated with the prices used in the regional comparisons (again, this is on average, some Basic Headings like Rice in Asia are less correlated in 2017).

Expression [14] implies that the regions can have a very uneven impact on the formation of inter-regional PPPs. Indeed, for the 2017 round, there were 5 separate regions in the ICP, participating in global linking: (i) Africa with 50 countries; (ii) Asia with 22 countries; (iii) the OECD with 49 countries; (iv) Latin America with 13 countries and (v) Western Asia with 12 countries, collecting the global core list of 631 products. Thus, the OECD and Africa region account for more than 70% of all weights, which is approximately proportional to the number of countries in the regions. More precisely, the regional weights are equal to the regional counts of collected prices multiplied by importance weights.

Starting from individual countries, we can summarize collected price data by region:

Table 1: Prices Collected in Individual Countries vs. Maximum Possible Prices to be Collected, grouped by Region in 2017 ICP

	All Regions	Africa	Asia	OECD	Latin America	Western Asia
Maximum # of Prices Possible	92126	31550	13882	30919	8203	7572
Actual # of Prices Collected	51332	17948	17501	7856	3239	4788
Percentage of Prices Collected (fill rate)	54.9	56.9	51.3	56.6	39.5	63.2
Regional share of prices collected	100.0	35.5	14.1	34.6	6.4	9.5
Average product weight	2.54	2.47	2.35	2.66	2.42	2.70
Regional share adj. for product weights	100.0	34.6	13.0	36.2	6.1	10.1

Now, because of the equivalence of system (9-10) and system (11-12), from now on we shall use system (9-10) and concentrate only on regional numbers for our analyses.

Table 2: Prices Collected in Individual Regions vs. Maximum Possible Prices to be Collected, by Region in 2017 ICP

	All Regions	Africa	Asia	OECD	Latin America	Western Asia
Maximum # of Prices Possible	3155	631	631	631	631	631
Actual # of Prices Collected	2138	415	461	407	415	440
Percentage of Prices Collected	67.8	65.8	73.1	64.5	65.8	67.8
Regional share of prices collected	100.0	19.4	21.6	19.0	19.4	20.6
Average product weight	60.0	106.9	36.2	114.3	18.9	29.4
Regional share adj. for product weights	100.0	34.6	13.0	36.2	6.1	10.1

However, the actual fill rate for the data used in calculation is somewhat higher as some items are not priced in any region and therefore are excluded from calculations. On top of that, some more items are effectively excluded from computing inter-regional PPPs as they are only priced in one region. Altogether, there are 50 items like that, so the actual number of items used in calculating the inter-regional PPPs is 581, and the overall fill rate increases to above 72% from 68.8%.

Table 3: Prices Collected in Individual Regions vs. Maximum Possible Prices to be Collected, by Region in 2017 ICP, including only effective products

	All Regions	Africa	Asia	OECD	Latin America	Western Asia
Maximum # of Prices Possible	2905	581	581	581	581	581
Actual # of Prices Collected	2097	409	448	394	413	433
Percentage of Prices Collected	72.2	70.4	77.1	67.8	71.1	74.5
Regional share of prices collected	100.0	19.5	21.4	18.8	19.7	20.6
Average product weight	60.1	107.1	36.6	114.7	18.9	29.5
Regional share adj. for product weights	100.0	34.8	13.0	35.9	6.2	10.1

The above table shows dominance of two regions in terms of weights, - Africa is responsible for 34.8% and OECD – for 35.9%. Thus, Africa has approximately six times higher weight than Latin America and the same goes for the OECD. As we shall see below the regional imbalance has a significant effect on the inter-regional PPPs.

At the same time, the regional price data fill rates range from 67.8% (OECD) to 77.1% (Asia), with the overall fill rate standing at 72.2%. This can be deemed acceptable for most Basic Headings. These fill rates for many BHs could probably be increased, but we have to be careful here lest it might happen at the cost of a possible loss of data quality as cross-country product comparability may decline and adding more work to their plate may strain the resources of the countries devoted to ICP even further. This is a classic example of the *less is more* principle: it's better to have less data but of higher quality than more noise.

However, some Basic Headings that do suffer from low fill rates can be significantly improved upon by incorporating more additional items into the Global Core List, pricing more existing items and, more importantly, using better intra- and inter-regional validation. On the other hand, another, and possibly more fruitful, solution for some of those problematic Basic Headings would be to use reference PPPs. This concept implies estimating parities of hard-to-measure BHs based on one or more other basic headings that had already been calculated. The concept of reference PPPs is widely used in the ICP. Thus, Household consumption includes 22 BHs estimated with reference PPPs (see Blades and Dikhanov (2013) [18Chapter17.pdf \(worldbank.org\)](#) for more detail).

3. Impact of large regions on the inter-regional linking

Under the current linking methodology, larger regions (Africa and OECD) have a disproportionate impact on the global comparison. The impact can be evaluated by removing one region at a time from the global calculations and comparing the results of the remaining four regions to the original 5-region calculations. Below are listed 10 BHs with the largest impact both for Africa and OECD. We can see that the impact reaches 91% in the case of Passenger transport by railway for the OECD, and 90% for Jewelry, clocks, and watches in Africa, which are very substantial numbers.

The deviations across Basic Headings (i.e., relative prices) will swing obviously more. Thus, removing Africa from linking causes PPP for Jewelry, clocks, and watches in Asia go up by 71.7% versus OECD, and PPP for Accommodation services go down by 43.1%. I.e., relative price between those two BHs more than triples as the result of removing Africa. Removal of smaller regions predictably affects global stability less, as they contain fewer countries.

Table 4. Impact of Africa on other regions in linking (OECD=100), CPDW

	ASI	EUO	LAC	WAS	Max/min
Jewelry, clocks, and watches	172%	100%	145%	190%	90%
Accommodation services	57%	100%	86%	101%	77%
Therapeutic appliances and equipment	73%	100%	103%	102%	40%
Carpets and other floor coverings	82%	100%	91%	113%	37%
Other meats and meat preparations	114%	100%	111%	137%	37%
Recreational and sporting services	120%	100%	93%	103%	29%
Motorcycles	127%	100%	124%	104%	27%
Other fuels	121%	100%	126%	118%	26%
Recording media	102%	100%	112%	93%	20%
Repair and hire of footwear	91%	100%	101%	84%	20%

Table 5. Impact of OECD other regions in linking (4 regions=100), CPDW

	AFR	ASI	LAC	WAS	Max/min
Passenger transport by railway	125%	124%	99%	65%	91%
Maintenance and repair of personal transport equipment	96%	104%	90%	112%	25%
Other services in respect of personal transport equipment	111%	102%	90%	98%	24%
Recording media	103%	112%	97%	90%	24%
Therapeutic appliances and equipment	97%	112%	102%	91%	23%
Carpets and other floor coverings	98%	95%	115%	93%	23%
Garden and pets	103%	91%	97%	110%	21%
Appliances, articles, and products for personal care	94%	102%	93%	113%	21%
Furniture and furnishings	99%	112%	93%	97%	20%
Rice	96%	113%	96%	97%	18%

Thus, it would be desirable indeed to try and limit these effects due to larger regions.

4. Bringing regional neutrality to linking with neutral CPDW (geometric IDB) index

The ICP has the fixity of regional results principle at its core. The principle dictates that results within any region are not affected by any other region's results. This creates a guarantee of stability for the intra-regional results within the global comparison. We don't have those requirements for the inter-regional linking, but it seems desirable to keep the global results as stable as possible.

The rationale behind it is quite straightforward. Consider a simplified case of N regions when the price matrix is full, and all items are of equal importance. In this case, adding any number M regions to the CPD regression on N+M regions does not change the inter-regional PPPs within the original N regions (i.e., when the CPD is applied to N regions only). In fact, this is true even in case when the additional M regions have an incomplete price matrix and arbitrary weights.

As we have seen in the previous chapter the current method of linking regions creates a serious imbalance where larger regions exert much bigger influence than others. Dikhanov (2022) suggested two ways to

resolve this problem. We will consider the first of the indices suggested (the geometric IDB index) as the second one is closely related to it and generates almost the same results.

From the above, the current linking method is written as follows:

$$(11) \rho_r = \sum_{n=1}^N w_n(r) [\ln p_n(r) - \beta_n] / \sum_{n=1}^N w_n(r); \quad r = 1, \dots, R;$$

$$(12) \beta_n = \sum_{r=1}^R w_n(r) [\ln p_n(r) - \rho_r] / \sum_{r=1}^R w_n(r); \quad n = 1, \dots, N.$$

where

$$(13) w_n(r) = \sum_{k=1}^{K(r)} w_{rkn}, \text{ or total weight of product } n \text{ in region } r;$$

$$(14) \ln p_n(r) = \sum_{k=1}^{K(r)} w_{rkn} \ln p_{rkn} / \sum_{k=1}^{K(r)} w_{rkn}, \text{ or log of weighted geomean price of product } n \text{ in region } r.$$

Let's introduce two new shares:

$$(15) s_n(r) = w_n(r) / \sum_{n=1}^N w_n(r), \text{ or share of product } n \text{ in region } r;$$

$$(16) v_n(r) = w_n(r) / \sum_{r=1}^R w_n(r), \text{ or share of region } r \text{ in global consumption of product } n.$$

Then system (11-12) can be re-written as:

$$(17) \rho_r = \sum_{n=1}^N s_n(r) [\ln p_n(r) - \beta_n]; \quad r = 1, \dots, R;$$

$$(18) \beta_n = \sum_{r=1}^R v_n(r) [\ln p_n(r) - \rho_r]; \quad n = 1, \dots, N.$$

The modification suggested by Dikhanov (2022) involves the following normalization of the weight system $\{w_{rkn}\}$:

$$(19) \tilde{w}_{rkn} = w_{rkn} / \sum_{k=1}^{K(r)} \sum_{n=1}^N w_{rkn},$$

This does not change $\ln p_n(r)$ expression, but regional weights $w_n(r)$ become:

$$(20) \tilde{w}_n(r) = w_n(r) / \sum_{k=1}^{K(r)} \sum_{n=1}^N w_{rkn} = w_n(r) / \sum_{n=1}^N w_n(r) = s_n(r)$$

And weights $\{v_n(r)\}$ become:

$$(21) \tilde{v}_n(r) = s_n(r) / \sum_{r=1}^R s_n(r)$$

Thus, the neutral CPDW system can be written as:

$$(22) \rho_r = \sum_{n=1}^N s_n(r) [\ln p_n(r) - \beta_n]; \quad r = 1, \dots, R;$$

$$(23) \beta_n = \sum_{r=1}^R \tilde{v}_n(r) [\ln p_n(r) - \rho_r]; \quad n = 1, \dots, N.$$

In this system the weighting scheme become neutral with each region exercising the same amount of influence. It can be shown that system (22-23) is the geometric version of the IDB index which is written as (see ANNEX):

$$(24) \pi_r = \sum_{n=1}^N s_n(r) [p_n(r) / \alpha_n]; \quad r = 1, \dots, R;$$

$$(25) \alpha_n = \sum_{r=1}^R \tilde{v}_n(r) [p_n(r) / \pi_r]; \quad n = 1, \dots, N.$$

The IDB index was designed specifically to remove the large country's bias from the GK, and the geometric version of it does the same for the CPDW linking method¹².

¹² The second index suggested by Dikhanov (2022) is closely related to the geometric IDB and is defined as:

$$(26) \bar{w}_{rkn} = w_{rkn} / \sum_{k=1}^{K(r)} N(r, k) = w_{rkn} / N(r)$$

As in the case of the geometric IDB, this does not change $\ln p_n(r)$ expression, but regional weights $w_n(r)$ become:

$$(27) \bar{w}_n(r) = w_n(r) / N(r)$$

Applying the geometric IDB does indeed bring significant reduction in the biases. The Tables below show significantly lower deviations for the same Basic Headings as in the previous Chapter.

Table 6. Impact of Africa on other regions in linking (OECD=100), neutral CPDW

	ASI	EUO	LAC	WAS	Max/min
Jewelry, clocks, and watches	132%	100%	117%	144%	44%
Accommodation services	71%	100%	92%	100%	41%
Therapeutic appliances and equipment	86%	100%	101%	101%	18%
Carpets and other floor coverings	92%	100%	96%	105%	14%
Other meats and meat preparations	108%	100%	106%	119%	19%
Recreational and sporting services	111%	100%	97%	103%	15%
Motorcycles	113%	100%	110%	103%	13%
Other fuels	108%	100%	112%	108%	12%
Recording media	101%	100%	108%	94%	15%
Repair and hire of footwear	97%	100%	102%	93%	10%

Table 7. Impact of OECD other regions in linking (4 regions=100), neutral CPDW

	AFR	ASI	LAC	WAS	Max/min
Passenger transport by railway	109%	108%	100%	85%	27%
Maintenance and repair of personal transport equipment	98%	102%	96%	105%	10%
Other services in respect of personal transport equipment	103%	102%	96%	99%	8%
Recording media	102%	107%	98%	94%	14%
Therapeutic appliances and equipment	98%	107%	102%	94%	15%
Carpets and other floor coverings	98%	97%	110%	96%	14%
Garden and pets	102%	95%	99%	105%	11%
Appliances, articles, and products for personal care	97%	101%	98%	104%	7%
Furniture and furnishings	99%	109%	94%	98%	16%
Rice	98%	108%	97%	98%	11%

Applying the neutral CPDW (or the geometric IDB) obviously changes the regional PPPs. The Table below shows 10 Basic Headings with largest deviations vs. the original CPDW.

Table 8. Effect of applying the neutral CPDW vs. the original CPDW on some Basic Headings

	AFR	ASI	EUO	LAC	WAS	max/min
Passenger transport by railway	111%	110%	103%	101%	78%	41%
Jewelry, clocks and watches	99%	108%	83%	103%	110%	32%
Accommodation services	100%	86%	107%	102%	107%	25%
Maintenance and repair of personal transport	99%	91%	111%	105%	96%	22%
Other fuels	99%	102%	91%	108%	101%	19%
Repair of audio-visual, photo- and inf. equipment	93%	102%	106%	97%	103%	15%
Recreational and sporting services	98%	109%	101%	95%	99%	15%
Other meats and meat preparations	103%	99%	94%	98%	107%	15%
Small electric household appliances	104%	104%	92%	99%	102%	13%
Garden and pets	106%	103%	97%	94%	100%	13%

Obviously, the effects described above do not depend on the number of countries in a region alone, but also depend on the quality of data in that region.

5. Case of Pharmaceuticals

Pharmaceuticals is the largest BH in terms of the number of products – it has 57 of them, or 9% of the total number of products in Household Consumption. It is much smaller though in terms of expenditure shares – on average it covers somewhat more than 1% of total expenditures.

Pharmaceuticals has a relatively high data fill rate – around 69%, with two regions – Asia and Western Asia having a 100% rate, and even the lowest rate region (OECD) having 18 products priced. However, Africa is dominating in terms of weights (almost a half of the world total).

This BH presents an interesting case study for different indices. We will use the neutral CPDW index (geometric IDB) defined by system (22-23) and the IDB index described by system (24-25). Both indices are size neutral. CPDW implies linear preferences and perfect product substitution¹³, whereas IDB implies Leontief preferences and no product substitution. As there is hardly any substitution possible for drugs of various types¹⁴, and any such substitution would indeed be highly undesirable health-wise, to put it mildly, we can expect here the maximum extent of the between-product substitution effect to be revealed by these two indices.

Regional price level indices for Pharmaceuticals

	AFR	ASI	EUO	LAC	WAS
cpdw neutral (geometric IDB)	0.45	0.99	1.00	1.22	0.92
idb	0.45	0.99	1.00	1.22	0.89

There is practically no difference in the results. In fact, there is much more difference between the original CPDW and the neutral CPDW (geometric IDB) than between the geometric IDB and the straight IDB indices.

	AFR	ASI	EUO	LAC	WAS
cpdw, original	0.46	1.03	1.00	1.27	0.96
cpdw neutral (geometric IDB)	0.45	0.99	1.00	1.22	0.92

¹³ The CPD indexes implicitly assume *linear preferences for the purchasers over the products in scope*. The CPD model is consistent with purchasers in a country maximizing the linear utility function, $f(q_1, \dots, q_N) \equiv \sum_{n=1}^N \alpha_n q_n$ (subject to budget constraints) where the marginal utility parameters α_n appear in equations (1) and the q_n are total market purchases of product n for the country under consideration (see section 5 in Diewert (2022b)). The underlying economic model implies that the products are *perfect substitutes* (after quality adjustment) and should be closely related. Most of the ICP Basic Headings (BHs) are quite homogenous, but there are exceptions: consider the BHs Jewelry, Clocks and Watches, or Pharmaceuticals.

¹⁴ Except for five drugs which come in generic and originator's versions.

This occurs due to the disproportionate weight of Africa in the original CPDW calculation which affects other regions (Africa's total weight being 4217 out of 9219 for the world), which now is neutralized with the geometric IDB.

The error terms $\{\varepsilon_{r,n}\}$ [last pane of the table] reveal some high inconsistencies for several items¹⁵. For example, in log terms, the error term for an item in Asia reaches plus 0.84, and in Western Asia – minus 0.85 for another one (it is $\exp(0.85) = 2.34$ times lower than the value predicted by the regression, in nominal terms). Given that those numbers are already smoothed out in calculation of regional averages and the sheer number of items priced in that basic Heading, these are rather high error terms.

¹⁵ Analysis of the error terms $\{\varepsilon_{r,n}\}$ is part and parcel of the standard ICP diagnostics as described in David Roberts (2013), in "ICP Book - Measuring the Real Size of the World Economy", Chapter 9: Validation of ICP Regional Prices and Basic Heading PPPs", World Bank [[10Chapter9.pdf \(worldbank.org\)](#)].

Pharmaceuticals	regional prices, US\$					regional weights					diagnostics (CPD residuals)				
	AFR	ASI	EUO	LAC	WAS	AFR	ASI	EUO	LAC	WAS	AFR	ASI	EUO	LAC	WAS
Acetaminophen/Paracetamol, 500 mg, Tablet/Capsule, Gener	0.22	0.74	1.30	0.91	0.88	131	42	106	22	34	-51%	-7%	52%	-10%	16%
Aciclovir/Acyclovir, 200 mg, Tablet/Capsule, Originator	4.15	10.46			8.61	114	20			31	-5%	9%			-4%
Aciclovir/Acyclovir, 200 mg, Tablet/Capsule, Generic	1.93	6.01			6.64	116	33			28	-29%	6%			23%
Amoxicillin, 250 mg, Tablet/Capsule, Generic	0.72	1.25		2.08	1.22	113	43		11	28	10%	-14%		13%	-9%
Atenolol, 50 mg, Tablet/Capsule, Originator	1.43	2.92			2.29	115	34			31	10%	3%			-14%
Captopril, 25 mg, Tablet/Capsule, Originator	1.09	2.89			2.04	113	8			28	-3%	16%			-12%
Ceftriaxone, vial with 1 gram of powder for injection, Original	4.76	18.72			18.20	110	15			31	-40%	18%			22%
Ciprofloxacin, 500 mg, Tablet/Capsule, Originator	3.18	12.97		21.74	4.62	119	24		10	29	-29%	33%		60%	-63%
Diclofenac, 50 mg, Tablet/Capsule, Originator	1.77	4.17	3.16	5.99	3.25	120	28	55	23	34	0%	7%	-17%	20%	-10%
Fluoxetine, 20 mg, Tablet/Capsule, Originator		18.74			11.30		15			26		22%			-22%
Glibenclamide, 5 mg, Tablet/Capsule, Generic	0.50	0.81			1.35	110	28			23	1%	-29%			28%
Losartan, 50 mg, Tablet/Capsule, Originator	2.70	8.48			7.68	115	27			31	-23%	13%			10%
Metformin, 500 mg, Tablet/Capsule, Generic	0.70	1.10	0.84		1.80	117	37	90		23	17%	-16%	-41%		40%
Nifedipine retard, 20 mg, Retard tablet, Originator		5.17			2.13		29			27		41%			-41%
Omeprazole, 20 mg, Tablet/Capsule, Originator	3.07	20.97		8.85	6.35	114	12		12	34	-30%	84%		-26%	-28%
Omeprazole, 20 mg, Tablet/Capsule, Generic	1.28	2.98	3.69	3.73	4.04	115	45	116	12	28	-17%	-11%	14%	-12%	27%
Ranitidine, 150 mg, Tablet/Capsule, Originator		6.10		7.68	4.50		17		14	31		8%		7%	-15%
Acetylsalicylic acid, 500 mg, Tablet/Capsule, Originator	0.89	1.72	2.62	1.55	0.70	106	14	95	18	22	25%	12%	57%	-23%	-71%
Acetylsalicylic acid, 500 mg, Tablet/Capsule, Generic	0.37	0.71		1.33	0.72	109	22		6	25	-2%	-15%		24%	-7%
Loratidine, 10 mg, Tablet/Capsule, Originator	2.19	6.27		9.32	3.48	117	22		16	30	-11%	16%		31%	-36%
Albendazole, 400 mg, Tablet/Capsule, Originator	12.51	14.90			22.97	107	13			22	24%	-37%			13%
Amlodipine, 5 mg, Tablet/Capsule, Originator		7.29	3.39	9.85	5.62		33	80	19	34		22%	-52%	28%	3%
Atorvastatin, 40 mg, Tablet/Capsule, Originator		19.23		22.42	12.38		21		17	31		15%		7%	-22%
Azithromycin, 500 mg, Tablet/Capsule, Originator		28.96			26.29		20			24		1%			-1%
Doxycycline, 100 mg, Tablet/Capsule, Generic		1.68			2.21		30			27		-17%			17%
Enalapril, 10 mg, Tablet/Capsule, Originator		3.59			3.81		17			22		-7%			7%
Furosemide, 40 mg, Tablet/Capsule, Originator		1.54			2.84		13			28		-34%			34%
Ibuprofen, 400 mg, Tablet/Capsule, Originator		2.53	3.15	3.80	1.41		21	109	17	29		2%	27%	19%	-49%
Metronidazole, 250 mg, Tablet/Capsule, Originator	1.28	1.48		4.71	0.75	105	11		6	24	40%	-24%		68%	-85%
Metronidazole, 250 mg, Tablet/Capsule, Generic	0.52	0.85		1.13	0.38	129	21		3	25	39%	9%		14%	-62%
Oral rehydration salts, WHO/UNICEF formulation, Oral Suspe	0.79	1.53		2.38	2.71	122	26		10	25	-12%	-24%		-4%	40%
Simvastatin, 20 mg, Tablet/Capsule, Originator	3.33	6.69			6.35	107	15			32	5%	-4%			-2%
Acetaminophen/Paracetamol, 500 mg, Tablet/Capsule, Origin	0.86	1.41	1.80	1.80	0.87	115	41	107	22	31	26%	-4%	24%	-3%	-44%
Albendazole, 400 mg, Tablet/Capsule, Generic	5.17	7.21			16.16	115	23			24	1%	-44%			43%
Amlodipine, 5 mg, Tablet/Capsule, Generic		2.58	1.66	1.74	3.26		38	111	11	27		18%	-22%	-45%	49%
Amoxicillin, 250 mg, Tablet/Capsule, Originator	1.15	3.18		4.28	1.94	103	24		11	28	-8%	15%		21%	-27%
Atenolol, 50 mg, Tablet/Capsule, Generic	0.80	1.48	1.45		1.88	123	43	62		26	4%	-12%	-11%		19%
Atorvastatin, 40 mg, Tablet/Capsule, Generic		14.38	4.42	6.50	8.58		23	91		8	26	66%	-49%	-38%	21%
Azithromycin, 500 mg, Tablet/Capsule, Generic		16.96			22.04		33			25		-17%			17%
Captopril, 25 mg, Tablet/Capsule, Generic	0.80	1.49			1.82	115	25			25	2%	-15%			12%
Ceftriaxone, vial with 1 gram of powder for injection, Generic	1.86	6.41			12.80	117	27			28	-55%	-10%			66%
Ciprofloxacin, 500 mg, Tablet/Capsule, Generic	1.54	3.18		3.52	2.47	115	35		10	29	13%	6%		-7%	-12%
Diazepam 5 mg, Tablet/Capsule, Generic	0.59	0.93	1.18	0.75	0.80	109	20	60	7	17	32%	-1%	25%	-47%	-9%
Diazepam 5 mg, Tablet/Capsule, Originator	1.17	1.74		3.11	1.45	92	10		10	14	23%	-15%		19%	-27%
Diclofenac, 50 mg, Tablet/Capsule, Generic	0.55	1.72	1.89	1.61	2.46	123	31	72	16	31	-33%	2%	15%	-29%	45%
Doxycycline, 100 mg, Tablet/Capsule, Originator		6.24			3.36		13			28		27%			-27%
Enalapril, 10 mg, Tablet/Capsule, Generic		2.28			2.97		27			26		-17%			17%
Fluoxetine, 20 mg, Tablet/Capsule, Generic		3.83	3.23		7.37		27	100		22		-20%	-33%		53%
Furosemide, 40 mg, Tablet/Capsule, Generic		0.85			2.62		20			26		-60%			60%
Glibenclamide, 5 mg, Tablet/Capsule, Originator	1.10	1.24			1.38	109	14			32	39%	-28%			-10%
Ibuprofen, 400 mg, Tablet/Capsule, Generic		1.23	2.20	1.57	1.28		35	105	19	29		-19%	43%	-18%	-7%
Loratidine, 10 mg, Tablet/Capsule, Generic	1.23	2.49	3.21	1.86	2.60	119	29	98	14	26	9%	1%	30%	-52%	12%
Losartan, 50 mg, Tablet/Capsule, Generic	1.79	4.05	2.61		3.97	117	38	103		31	7%	10%	-31%		15%
Metformin, 500 mg, Tablet/Capsule, Originator	1.12	2.28			1.70	110	32			23	12%	5%			-17%
Nifedipine retard, 20 mg, Retard tablet, Generic		1.76			1.39		28			19		8%			-8%
Ranitidine, 150 mg, Tablet/Capsule, Generic		1.82		2.11	2.39		29		11	29		-8%		-18%	26%
Simvastatin, 20 mg, Tablet/Capsule, Generic	1.82	4.29	2.65		3.69	111	26	95		26	8%	15%	-30%		7%
Price level index	0.46	1.03	1.00	1.27	0.96					s.d.	24%	24%	35%	30%	33%
count	37	57	18	27	57	4217	1447	1655	355	1545					
fillrate	65%	100%	32%	47%	100%										

5. Case of Rice

Rice is on the larger side in terms of the number of products as far as the BHs go: it has 9 products (nominally, 11 products are listed but 2 of them are not priced by any region)¹⁶. It is quite an important BH for many countries, though.

Rice has about average price data fill rate – around 60%, with two regions – Africa and Asia - having close to 80-90% rates; however, Western Asia is pricing 3 items only. Again, Africa is dominant in terms of weights (almost a half of the world total).

This BH is probably one of the most homogenous and it is another interesting case study. Again, we will use the neutral CPDW index (geometric IDB) and the IDB to test the effects of different assumptions for underlying consumer preferences. Opposite to the case of Pharmaceuticals, all products in Rice BH can be considered substitutes, and we can expect here the maximum extent of the between-product substitution effect to be revealed by these two indices, just in the opposite direction from the case of Pharmaceuticals.

Regional price level indices for Rice

	AFR	ASI	EUO	LAC	WAS
cpdw neutral (geometric IDB)	0.62	0.68	1.00	0.73	1.00
idb	0.63	0.70	1.00	0.75	1.01

Again, there isn't any meaningful difference in the results. At the same time, for LAC the original CPDW is more different from the neutral CPDW (geometric IDB), with little difference to show for other regions:

	AFR	ASI	EUO	LAC	WAS
cpdw, original	0.63	0.67	1.00	0.78	1.00
cpdw neutral (geometric IDB)	0.62	0.68	1.00	0.73	1.00

The error terms $\{\varepsilon_{r,n}\}$ reveal some high inconsistencies for several items, reaching plus 0.57 (Basmati rice in LAC) and minus 0.46 (Long-grain rice, not parboiled, in OECD). However, in the case of LAC, the item has relatively small weight, so it does not affect the PPPs in a significant way. On average the error terms for this BH are about average.

It is tempting to ask countries to price more items. However, consumer patterns can differ by region and by country within a region significantly. For instance, Basmati rice in LAC is priced by 4 countries out of 13, and even there it is not normal to consume it, so it is assigned a weight of 1 in those countries, resulting in a total weight of 4 in LAC for that item. The value of the cpd residual $\varepsilon_{r,n}$ for Basmati rice in LAC is 0.57 which indicates a much higher price than predicted by the model. However, given that it is consumed in those countries by small immigrant or expatriate communities, or used as an add-on for cooking and not as

¹⁶ Actual computations will involve seven products, as one product is priced in Asia only and one in Africa only, and thus they have no effect on the multilateral PPPs. Therefore, the effective fill rate will become 86%, 100%, 57%, 71%, 43% in Africa, Asia, OECD, Latin America and Western Asia, respectively; with the overall rate for the world reaching 71%.

7. Case of Passenger transport by railway

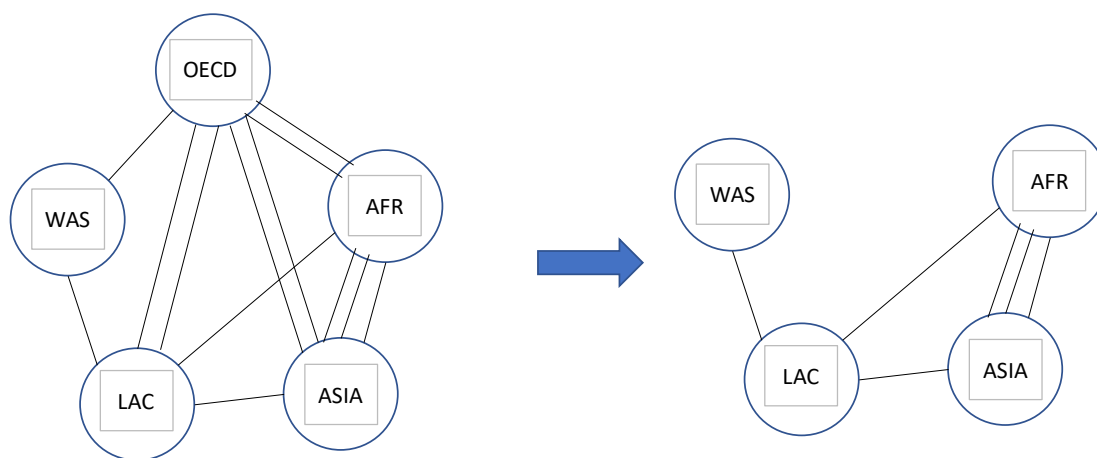
Passenger transport by railway is the worst behaving BH in terms of the single region effect. It contains 5 products. However, the number of effective products is four, as one product [Urban tram (rail) or tube, monthly ticket¹⁷] is priced by the OECD only and thus has no effect on inter-regional links. The overall fill rate (once accounted for one less product) is quite low - 60%, with three regions having 75%, LAC – 50%, and WAS – 25% (only one product is priced there). But the biggest problem is that WAS has direct links through that single product with only two other regions out of four possible: LAC and OECD.

Passenger transport by railway	regional prices, US\$					regional weights				
	AFR	ASI	EUO	LAC	WAS	AFR	ASI	EUO	LAC	WAS
Interurban transport, single ticket, 50 km	1.78	2.62	10.21	1.54		86	30	106	12	
Interurban transport, single ticket, 150 km	4.70	11.00				85	30			
Interurban transport, single ticket, 250 km	7.41	15.33	32.77			80	34	84		
Urban tram (rail) or tube, single ticket			2.74	0.99	0.52			71	9	9
Urban tram (rail) or tube, monthly ticket			73.89					89		
count/ weight	3	3	3	2	1	251	94	261	21	9
fill rate	75%	75%	75%	50%	25%					
weight						39%	15%	41%	3%	1%

As this BH was listed in Table 5 as the most problematic, let us see what happens when we remove OECD from the comparison. Then WAS is left with a single link to LAC, and due to its low weight and hence little impact, the position of WAS vis-à-vis other regions changes drastically when going from the 5-region linking to the 4-region one. The graph below depicts changes in links, with the number of lines representing the strength of the links (this is # of common products).

¹⁷ This product is available mostly in OECD and CIS regions.

Connected graphs for the 5- and 4-region linking



The current CPDW linking method is found to be quite problematic for this BH, as the max/min differences for the index, when going from the 5- to 4-country linking, reach 91% for the AFR vs. WAS pair (see the Table below). However, both the CPDW neutral (geometric IDB) and regular IDB are much more resilient: the max/min differences for the indices stand at 27% and 29%, respectively. It is also notable that those two indices produce almost identical results.

Price level index (4-region geomean = 1.00)

5-region linking	AFR	ASI	LAC	WAS		
cpdw, original	0.83	1.60	0.97	0.77		
cpdw neutral (geometric IDB)	0.93	1.78	0.99	0.61		
idb	0.92	1.77	0.99	0.62		
4-region linking (no OECD)						
cpdw, original	1.04	1.99	0.96	0.51		
cpdw neutral (geometric IDB)	1.01	1.93	0.99	0.52		
idb	1.00	1.92	0.99	0.52		
4-region vs. 5-region (ratio)					<i>max/min</i>	<i>s.d.</i>
cpdw, original	1.25	1.24	0.98	0.65	91%	24.3%
cpdw neutral (geometric IDB)	1.09	1.08	1.00	0.85	27%	9.4%
idb	1.09	1.09	1.00	0.85	29%	9.9%

Thus, the move to a neutral index such as the geometric IDB should improve the linking. In addition, more steps can be undertaken in order to improve results in this Basic Heading:

1. Require two effective products (i.e., products overlapping with at least one other region) as a minimum. WAS should be able to price at least one more item as, for example, Egypt, Tunisia, Morocco, Sudan all have railway systems. At the same time, it must be noted that the share of this BH in total expenditures is quite low in WAS region, thus any impact at the Household Consumption level will be rather limited.

2. Improve comparability of the items across countries and regions. Quality of the ride may be not quite comparable across the world, as such things as comfort, timeliness, state of the rolling stock may well not be fully taken into account. This incomparability leads to lower PPPs of other regions vs. OECD.

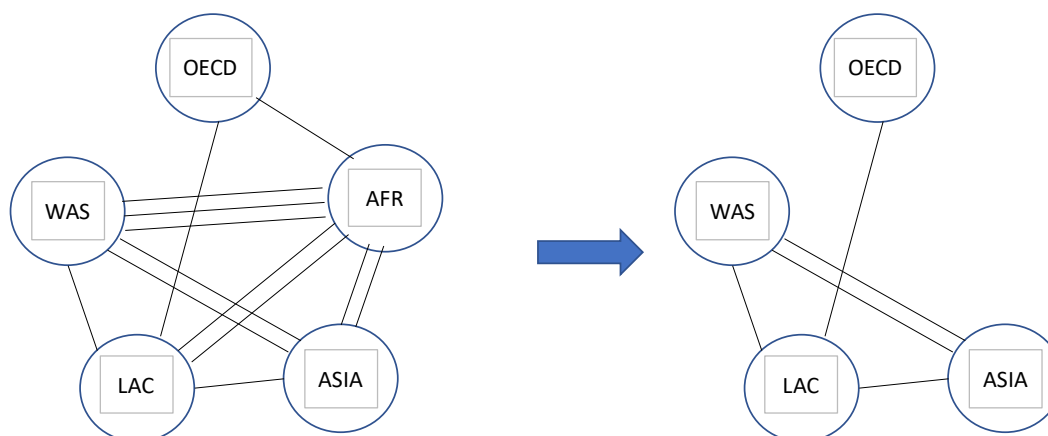
8. Case of Jewelry, Clocks and Watches

Jewelry, Clocks and Watches is another example of a poorly behaving BH in terms of the single region effect (see Table 4). It contains 5 products. However, the number of effective products is four, as one product [Wrist-watch, children's, SWATCH Flik Flak] is priced by the OECD only and thus has no effect on the inter-regional links. The overall fill rate (once accounted for one less product) is quite low - 60%, with Africa having 100%, WAS – 75%, Asia and LAC – 50%, and OECD – 25% (only one overlapping product is priced there). As with Passenger transport by railway, a major problem is that OECD has direct links through that single product with only two other regions out of four possible: LAC and Africa.

Jewelry, Clocks and Watches	regional prices, US\$					regional weights				
	AFR	ASI	EUO	LAC	WAS	AFR	ASI	EUO	LAC	WAS
Wristwatch, children's, SWATCH Flik Flak			52.90					91		
Wristwatch, men's, CITIZEN Eco-Drive BM6060	221.04	183.32		127.60	130.68	23	21		15	25
Analog travel alarm, quartz, BL	11.44				7.07	108				28
Wedding ring, 14 Karat gold, BNR	62.10		75.07	93.17		106		97	11	
Wall clock, SEIKO	22.17	30.52			22.52	109	27			30
Price level index	0.86	1.03	1.00	0.86	0.68					
count/ weight	4	2	1	2	3	346	48	97	26	83
fill rate	100%	50%	25%	50%	75%					
weight						58%	8%	16%	4%	14%

Thus, if we remove Africa from the comparison, OECD is left with a single link to LAC. In addition, in the 4-region setting WAS loses one product as nobody prices it anymore. The graph below depicts changes in links, with the number of lines representing the strength of the links (this is # of common products). In the 4-region setting the connection graph becomes quite minimalistic.

Connected graphs for the 5- and 4-region linking



The current CPDW linking method is found extremely problematic for this BH, as the max/min differences for the index, when going from the 5- to 4-country linking, reach 90% for the AFR vs. WAS pair (see Table below). However, both the CPDW neutral (geometric IDB) and regular IDB are more resilient: the max/min differences for the indices stand at 44% and 46%, respectively. Again, it is also notable that those two indices produce almost identical results.

Price level index (OECD = 1.00)

5-region linking		ASI	EUO	LAC	WAS		
cpdw, original		1.03	1.00	0.86	0.68		
cpdw neutral (geometric IDB)		1.34	1.00	1.06	0.89		
idb		1.33	1.00	1.05	0.88		
4-region linking (no Africa)							
cpdw, original		1.76	1.00	1.24	1.28		
cpdw neutral (geometric IDB)		1.77	1.00	1.24	1.29		
idb		1.77	1.00	1.24	1.29		
4-region vs. 5-region (ratio)						max/min	s.d.
cpdw, original		1.72	1.00	1.45	1.90	90%	33.8%
cpdw neutral (geometric IDB)		1.32	1.00	1.17	1.44	44%	16.5%
idb		1.33	1.00	1.18	1.46	46%	17.1%

Again, the move to a neutral index such as the geometric IDB improves the linking. However, given that the product definitions in this Basic Heading are quite loose (the watches are often substituted by different

models, travel alarm is described quite loosely), and the gold prices are especially suspect¹⁸, we are afraid that getting more prices may not improve the results in any significant way, and our recommendation for Jewelry, Clocks and Watches would be to switch to a reference PPP for inter-regional linking, as it is done already in some regions for intra-regional PPP estimation.

9. Recommendations

The current interregional linking method – the CPDW - is far from being satisfactory due to it being adversely affected by larger regions, such as Africa and OECD. The distortions due to the large-region bias could reach 90% in some cases. In addition, some Basic Headings suffer from the lack of product overlap which, in turn, is exacerbated by the current CPDW linking method.

To improve on the index number, a neutral version of the CPDW – the geometric IDB – is proposed. That index significantly improves stability of the linking and reduces the large-region bias. The index is block-decomposable meaning that it is possible to use it only with weighted regional price geomeans and regional weights as inputs, and thus it does not need utilizing the individual 146 countries' data in the regression.

Even though the regional price data fill rates range from 67.8% (OECD) to 77.1% (Asia), for some Basic Headings the rates can be much lower. Those Basic Headings can be improved upon by incorporating additional items into the Global Core List, pricing more existing items, but, more importantly, using better intra- and inter-regional validation. However, we have to be careful here and find a proper balance lest to overburden the countries, as adding more items to collect to their workload may strain their resources devoted to ICP even further, which could in turn affect data quality.

We also recommend having as least two effective products per Basic Heading (i.e., products overlapping with at least one other region). Another recommendation is to improve comparability of the items across countries and regions. For example, in public transportation, quality of the ride may be not quite comparable across the world, as such things as comfort, timeliness, state of the rolling stock may well not be fully taken into account.

On the other hand, a possibly more fruitful solution for some of those problematic Basic Headings could be to use reference PPPs. This concept implies estimating parities of hard-to-measure BHs based on one or more other basic headings that had already been calculated. The concept of reference PPPs is widely used in the ICP.

¹⁸ For example, in countries of one of the ICP regions the US\$ prices for individual products are reported within the following ranges:

	min	max
Wristwatch, men's, CITIZEN Eco-Drive BM6060	85.0	394.8
Analog travel alarm, quartz, BL	3.1	18.6
Wedding ring, 14 Karat gold, BNR	13.0	138.2
Wall clock, SEIKO	6.0	58.3

References

1. Diewert, W.E. (2022), “*The Chain Drift Problem and Multilateral Indexes*”, Chapter 7 in *Consumer Price Index Theory*, Washington D.C.: International Monetary Fund, <https://www.imf.org/en/Data/Statistics/cpi-manual>.
2. Diewert, W.W. (2008), *New Methodological Developments for the International Comparison Program*, Paper presented at the Joint UNECE/ILO Meeting on Consumer Prices Indices, May 8-9, Geneva.
3. Diewert, W.W. (2008), *New Methodology for Linking the Regions*, Discussion Paper 08-07, Department of Economics, The University of British Columbia.
4. Dikhanov (2022), *Finetuning ICP Linking Procedure*, ICP TAG presentation, November 2022.
5. Dikhanov (1996), “*Sensitivity of PPP-Based Income Estimates to Choice of Aggregation Procedures*”, World Bank, <https://thedocs.worldbank.org/en/doc/4dd69ce0c60b9d4a3dcc59886292e002-0050022021/original/ICPPAPER-total.pdf>
6. Reference PPPs, Chapter 17, *ICP Book*, World Bank, 2013, [Chapter17.pdf - worldbank.org](#).

ANNEX I: Geometric GK and IDB vs. regular GK and IDB

The weight system introduced for ICP 2011 as system (9-10) implies that the importance weights w_{rkn} are commensurable across countries and across products¹⁹, i.e., those weights are stated in real terms and, thus, could be interpreted as world price of product n α_n multiplied by quantity q_{rkn} , i.e., $w_{rkn} = \alpha_n q_{rkn}$. For regions, the constituting country weights aggregate into $w_n(r) = \alpha_n q_n(r)$, where $q_n(r) = \sum_{k=1}^{K(r)} q_{rkn}$. From the above, $p_n(r) = \exp(\sum_{k=1}^{K(r)} w_{rkn} \ln p_{rkn} / \sum_{k=1}^{K(r)} w_{rkn})$.

GK system

The GK system for regions can be written as:

$$(26) \pi_r = (\sum_{n=1}^N s_n^*(r) [\alpha_n / p_n(r)])^{-1}; \quad r = 1, \dots, R;$$

$$(27) \alpha_n = \sum_{r=1}^R v_n(r) [p_n(r) / \pi_r]; \quad n = 1, \dots, N.$$

where

$$(28) s_n^*(r) = p_n(r) q_n(r) / \sum_{n=1}^N p_n(r) q_n(r), \text{ or nominal expenditure share of product } n \text{ in region } r;$$

$$(29) v_n(r) = q_n(r) / \sum_{r=1}^R q_n(r) = w_n(r) / \sum_{r=1}^R w_n(r), \text{ or share of region } r \text{ in global consumption of product } n.$$

Noting that

$$\pi_r = \sum_{n=1}^N p_n(r) q_n(r) / \sum_{n=1}^N \alpha_n q_n(r) = \sum_{n=1}^N [p_n(r) / \alpha_n] [\alpha_n q_n(r)] / \sum_{n=1}^N \alpha_n q_n(r) = \sum_{n=1}^N [p_n(r) / \alpha_n] w_n(r) / \sum_{n=1}^N w_n(r) = \sum_{n=1}^N s_n(r) [p_n(r) / \alpha_n],$$

$$\text{where } s_n(r) = w_n(r) / \sum_{n=1}^N w_n(r).$$

Thus, the GK system can be re-written as:

$$(30) \pi_r = \sum_{n=1}^N s_n(r) [p_n(r) / \alpha_n]; \quad r = 1, \dots, R;$$

$$(31) \alpha_n = \sum_{r=1}^R v_n(r) [p_n(r) / \pi_r]; \quad n = 1, \dots, N.$$

And the geometric version of the GK will be the CPDW:

$$(32) \rho_r = \sum_{n=1}^N s_n(r) [\ln p_n(r) - \beta_n]; \quad r = 1, \dots, R;$$

$$(33) \beta_n = \sum_{r=1}^R v_n(r) [\ln p_n(r) - \rho_r]; \quad n = 1, \dots, N.$$

IDB system

The IDB system for regions can be written as:

$$(34) \pi_r = (\sum_{n=1}^N s_n^*(r) [\alpha_n / p_n(r)])^{-1}; \quad r = 1, \dots, R;$$

$$(35) \alpha_n = (\sum_{r=1}^R \tilde{v}_n^*(r) [\pi_r / p_n(r)])^{-1}; \quad n = 1, \dots, N.$$

where

$$(36) s_n^*(r) = p_n(r) q_n(r) / \sum_{n=1}^N p_n(r) q_n(r);$$

$$(37) \tilde{v}_n^*(r) = s_n^*(r) / \sum_{r=1}^R s_n^*(r).$$

¹⁹ The importance weights $\{3; 1\}$ were introduced in order to “reward” countries with more and of higher relevance (“importance”) price data to have a larger effect on the formation of the regional price level. However, this backfired at the global level where larger regions commanded the lion’s share of influence in determining the regional PPPs. Hence the need for neutrality in regional linking.

From Dikhanov (1996), p.6²⁰ it follows that:

$$(38) \alpha_n = (\sum_{r=1}^R \tilde{v}_n^*(r) [\pi_r / p_n(r)])^{-1} = \sum_{r=1}^R \tilde{v}_n(r) [p_n(r) / \pi_r],$$

where $\tilde{v}_n(r) = s_n(r) / \sum_{r=1}^R s_n(r)$, or that the world price for product n can be equivalently expressed via the harmonic mean with nominal shares, or the arithmetic mean with real shares.

Thus, using this result and expression (30), we can write the IDB system as:

(39) $\pi_r = \sum_{n=1}^N s_n(r) [p_n(r) / \alpha_n];$	$r = 1, \dots, R;$
(40) $\alpha_n = \sum_{r=1}^R \tilde{v}_n(r) [p_n(r) / \pi_r];$	$n = 1, \dots, N.$

And the geometric version of IDB (the neutral CPDW) system is written as:

(22) $\rho_r = \sum_{n=1}^N s_n(r) [\ln p_n(r) - \beta_n];$	$r = 1, \dots, R;$
(23) $\beta_n = \sum_{r=1}^R \tilde{v}_n(r) [\ln p_n(r) - \rho_r];$	$n = 1, \dots, N.$

²⁰ Dikhanov (1996), “Sensitivity of PPP-Based Income Estimates to Choice of Aggregation Procedures”, World Bank, <https://thedocs.worldbank.org/en/doc/4dd69ce0c60b9d4a3dcc59886292e002-0050022021/original/ICPPAPER-total.pdf>

ANNEX II. COMPARATIVE TABLES FOR DIEWERT, DIKHANOV and SERGEEV LINKING METHODS

First, consider the general weighted case:

Table 1. Weighted case

	Diewert (the current method)	Dikhanov	Sergeev
Inputs	[1] p_{rkn} is the price of product n in region r , country k , adjusted by intra-regional BH PPP; [2] w_{rkn} is the weight of product n in region r , country k .		
Direct one-stage computation	System (9-10): $\rho_r = \sum_{n=1}^N \sum_{k=1}^{K(r)} w_{rkn} [\ln p_{rkn} - \beta_n] / \sum_{n=1}^N \sum_{k=1}^{K(r)} w_{rkn}$; $\beta_n = \sum_{r=1}^R \sum_{k=1}^{K(r)} w_{rkn} [\ln p_{rkn} - \rho_r] / \sum_{r=1}^R \sum_{k=1}^{K(r)} w_{rkn}$;	Trasformation (19) to achieve regional neutrality is applied to weights [2]: $w^*_{rkn} = w_{rkn} / \sum_{n=1}^N \sum_{k=1}^{K(r)} w_{rkn}$	Does not exists, it is a two-stage process
		System (9-10) becomes: $\rho_r = \sum_{n=1}^N \sum_{k=1}^{K(r)} w^*_{rkn} [\ln p_{rkn} - \beta_n] / \sum_{n=1}^N \sum_{k=1}^{K(r)} w^*_{rkn} = \sum_{n=1}^N \sum_{k=1}^{K(r)} w_{rkn} [\ln p_{rkn} - \beta_n] / \sum_{n=1}^N \sum_{k=1}^{K(r)} w_{rkn}$ $\beta_n = \sum_{r=1}^R \sum_{k=1}^{K(r)} w^*_{rkn} [\ln p_{rkn} - \rho_r] / \sum_{r=1}^R \sum_{k=1}^{K(r)} w^*_{rkn}$;	
Regional prices and weights	Implicit (not needed in calculations, but simplifies the setup)		Explicit Regional prices are explicitly generated as weighted geomeans; Regional weights are postulated ad hoc: depends on the majority weight
	(14) $\ln p_n(r) = \sum_{k=1}^{K(r)} w_{rkn} \ln p_{rkn} / \sum_{k=1}^{K(r)} w_{rkn}$		
	(13) $w_n(r) = \sum_{k=1}^{K(r)} w_{rkn}$	$w^*_n(r) = \sum_{k=1}^{K(r)} w_{rkn} / \sum_{n=1}^N \sum_{k=1}^{K(r)} w_{rkn}$	$Q_n(r) = \{1; 3\}$, if majority of prices have weight 3, then regional weight is 3, otherwise it is 1.
Calculation using Regional prices and weights (not needed, but possible for Diewert & Dikhanov methods)	Calculated using CPDW on the regional prices and regional weights		
	System (9-10) collapses into system (11-12): $\rho_r = \sum_{n=1}^N w_n(r) [\ln p_n(r) - \beta_n] / \sum_{n=1}^N w_n(r)$; $\beta_n = \sum_{r=1}^R w_n(r) [\ln p_n(r) - \rho_r] / \sum_{r=1}^R w_n(r)$; Or, equivalently, into system (32-33) : $\rho_r = \sum_{n=1}^N s_n(r) [\ln p_n(r) - \beta_n]$; $\beta_n = \sum_{r=1}^R v_n(r) [\ln p_n(r) - \rho_r]$; where: $v_n(r) = w_n(r) / \sum_{r=1}^R w_n(r)$	System (11-12 and, correspondingly, 32-33) becomes then system (22-23) : $\rho_r = \sum_{n=1}^N s_n(r) [\ln p_n(r) - \beta_n]$; $\beta_n = \sum_{r=1}^R v^*_n(r) [\ln p_n(r) - \rho_r]$; where: $v^*_n(r) = s_n(r) / \sum_{r=1}^R s_n(r)$	$\rho_r = \sum_{n=1}^N Q_n(r) [\ln p_n(r) - \beta_n] / \sum_{n=1}^N Q_n(r)$; $\beta_n = \sum_{r=1}^R Q_n(r) [\ln p_n(r) - \rho_r] / \sum_{r=1}^R Q_n(r)$
Specifics	One consistent stage of calculations		Two unrelated stages
Regional neutrality	Not neutral	Neutral	Neutral

As we can see, the difference between Diewert and Dikhanov methods in the general weighted case boils down to using weights w^*_{rkn} instead of w_{rkn} , which ensures regional neutrality and normalizes regional weights. The Dikhanov method is closely related to the Diewert method, and thus can be easily explained as a modification of the current methodology to remove the regional bias. Diewert and Dikhanov methods will become identical if the summary weight of countries per region were the same. Sergeev method will be close to Dikhanov method if mixes of products and their importance were approximately balanced among the regions.

Next, let's derive the **unweighted** case from the weighted case:

Table 2. Unweighted case

	Diewert (the 2005 method)	Dikhanov	Sergeev
Inputs	[1] p_{rkn} is the price of product n in region r , country k , adjusted by intra-regional BH PPP; [2] w_{rkn} is the weight of product n in region r , country k , in the unweighted case it is $\{0,1\}$		
Notation	$N(r,k)$ is the number of products priced in country k of region r ; $N^*(r,n)$ is the number of countries in region r that priced product n ; $N(r)$ is the number of products priced by all countries in region r ; $N^*(n)$ is the number of countries that priced product n ; $s_n(r) = N^*(r, n) / N(r)$; $v_n(r) = N^*(r, n) / N^*(n)$; $v_n^*(r) = s_n(r) / \sum_{r=1}^R s_n(r)$.		
Direct one-stage computation	System (7c-8c): $\rho_r = \sum_{n \in S(r,k)} \sum_{k=1}^{K(r)} [\ln p_{rkn} - \beta_n] / N(r)$; $\beta_n = \sum_{r=1}^R \sum_{k \in S^*(r,n)} [\ln p_{rkn} - \rho_r] / N^*(n)$ $= \sum_{r=1}^R v_n(r) \sum_{k \in S^*(r,n)} [\ln p_{rkn} - \rho_r] / N^*(r,n)$	Trasformation to achieve regional neutrality is applied to weights [2]: $w_{rkn}^* = w_{rkn} / N(r)$, or $\{0; 1/N(r)\}$ instead of $\{0,1\}$	Does not exists, it is a two-stage process
		System (7c-8c) transforms into: [3] $\rho_r = \sum_{n \in S(r,k)} \sum_{k=1}^{K(r)} [\ln p_{rkn}^* - \beta_n] / N(r)$; (same as Diewert) [4] $\beta_n = \sum_{r=1}^R [N^*(r,n) / N(r)] \sum_{k \in S^*(r,n)} [\ln p_{rkn}^* - \rho_r] / \sum_{r=1}^R s_n(r) \sum_{k \in S^*(r,n)} [\ln p_{rkn}^* - \rho_r] / N^*(r,n) =$ $= \sum_{r=1}^R s_n(r) \sum_{k \in S^*(r,n)} [\ln p_{rkn}^* - \rho_r] / N^*(r,n) =$ $= \sum_{r=1}^R v_n^*(r) \sum_{k \in S^*(r,n)} [\ln p_{rkn}^* - \rho_r] / N^*(r,n)$	
Regional prices and weights	Implicit (not needed in calculations, but simplifies the setup)		Explicit Regional prices are explicitly generated as unweighted geomeans; Regional weights are one.
	$\ln p_n(r) = \sum_{k=1}^{K(r)} \ln p_{rkn} / N^*(r,n)$		
	$w_n(r) = N^*(r, n)$	$w_n^*(r) = N^*(r, n) / N(r)$	$Q_n(r) = 1$
Calculation using Regional prices and weights (optional for Diewert & Dikhanov methods, necessary for Sergeev)	Calculated using CPDW on the regional prices and regional weights		Calculated using CPD on the regional prices
	System (7c-8c) collapses into system: $\rho_r = \sum_{n=1}^N N^*(r, n) [\ln p_n(r) - \beta_n] / N(r)$; $\beta_n = \sum_{r=1}^R N^*(r, n) [\ln p_n(r) - \rho_r] / N^*(n)$; Or, equivalently, into system: $\rho_r = \sum_{n=1}^N s_n(r) [\ln p_n(r) - \beta_n]$; $\beta_n = \sum_{r=1}^R v_n(r) [\ln p_n(r) - \rho_r]$	System [3-4] transforms into: $\rho_r = \sum_{n=1}^N s_n(r) [\ln p_n(r) - \beta_n]$; $\beta_n = \sum_{r=1}^R v_n^*(r) [\ln p_n(r) - \rho_r]$	$\rho_r = \sum_{n=1}^N [\ln p_n(r) - \beta_n] / M(r)$; $\beta_n = \sum_{r=1}^R [\ln p_n(r) - \rho_r] / M^*(n)$; where, $M(r)$ is the number of products priced in region r ; $M^*(n)$ is the number of regions that priced product n . [<i>note the difference between $N(r)$, $N^*(n)$, and $M(r)$, $M^*(n)$</i>]
Specifics	One consistent stage of calculations		Two unrelated stages
Regional neutrality	Not neutral	Neutral	Neutral

Again, the difference between Diewert and Dikhanov methods in the unweighted case boils down to using weights $\{0; 1/N(r)\}$ instead of $\{0,1\}$, which ensures regional neutrality. The Dikhanov method is closely related to the Diewert method, and thus can be easily explained as a modification of it to remove the regional bias. Diewert and Dikhanov methods will become identical if the number of countries per region were the same. Sergeev method will be close to Dikhanov method if mixes of products are approximately balanced among the regions.