

# International Comparison Program

---

[07.02]

## Linking the Regions

*Paul Konijn & Francette Koechlin*



**5<sup>th</sup> Technical Advisory Group Meeting**

**April 18-19, 2011**

**Washington DC**

## Linking the regions

### Linking at BH level (consumer goods only)

Starting point:

- matrix of prices of 600 products for 110 BHs by 180 countries
- product prices are marked for importance
- matrices of regional BH PPPs and expenditures for each region

Methodology agreed at previous TAG meeting:

- same as for Ring comparison of 2005, with inclusion of "importance" in regression
- conversion of all prices to regional numeraire by dividing by regional PPPs for BH
- RPID method: Region-Product-Importance-Dummies to calculate inter-regional PPPs for each BH

RPID, in brief, views each region as a single country. The national average prices for a region, expressed in the regional numeraire, can be seen as different observations for the product for this region. However, it is not clear how the Importance dimension is to be interpreted in the RPID method – see below.

Problems:

1. The matrix of prices is sparse, in both dimensions:
  - a. For BHs: sometimes there are only a few (or even one) product(s) per BH -> this may lead to unreliable BH PPPs
  - b. For countries: in some regions only a few countries may price a product -> this may lead to biased BH PPPs if the few countries within a region that priced the product are not representative for the region as a whole (like it was the case in the 2005 Ring comparison).

Regarding a): Should we group BHs together and perform the RPID on these groups instead? Subsequently, the group linking factors can be applied to each of the underlying BHs. We could for example set a minimum number of products (5? 10?) to be included in one regression.

Regarding b): to prevent the regressions to be dominated by a few countries within individual regions, we can also try to define some minimum standards for the inclusion of products in the RPID regression. For example:

- one rule could be that all prices entering the RPID must also have been used for the regional PPPs.
- In addition, one could require that at least, say, 25% (or should it be higher? lower? we don't want to throw away too many prices) of the countries within each region must have priced the product before this product should be included in the RPID as prices for that region.

2. Regions with many countries dominate the RPID regression. Hence, Eurostat/OECD may have 40 observations, Latin America 10, Africa 45, etc.

The alternative proposed by Sergey is to first calculate regional average prices and regional average importance, so that each region has only one price in the RPID. Then, each region would be treated symmetrically.

The question is therefore should each region or each country be treated symmetrically? One could argue that this is a global regression, and therefore bigger regions should have a bigger impact.

In Sergey's method it is still very important that there are a sufficient number of countries pricing the product for each region in order to obtain reliable and unbiased average regional prices.

3. Unclear interpretation of the Importance dummy in the RPID model.

Formula (1) in the ICP manual, chapter 14, only defines one representativity dummy per country, whereas representativity depends on *product and country*. In fact, formula (1) seems to "elevate" representativity to the BH level, whereas it is a concept that applies to products within the BH. Paragraph 28 of the same chapter also looks at the number of asterisk as a share of prices within the BH instead of within the product – it does not seem correct.

There seems to be three ways out of this (to be investigated further):

- Multiply the number of dummy variables for importance: there should be  $(n-1)$ , where  $n$ =number of products.
- Drop the importance variable from the regression (i.e. treat all products as important) and include only prices that are important (i.e. delete all the non-important prices). This would of course reduce the number of prices and would aggravate the problems under 1 (sparsity).
- Apply Sergey's method mentioned under 2.

### **Linking above the BH level (all BHs)**

Starting point:

- matrix of BH PPPs and expenditures: 150 BHs by 180 countries – in principle full matrix (no gaps)

Two methods were still on the table after the previous TAG, both involve the calculation of EKS-PPPs in a global calculation, i.e. including all countries.

1. Heston/Dikhanov etc: linking via the volumes, i.e. maintaining fixity of the inter-regional volume shares. Comes down to calculating scaling factors for each region as weighted harmonic means of the ratios between the PPPs derived in the global calculation and the regional PPPs. Can be shown to minimize the changes to be made to the world volume shares to maintain fixity.

2. Eurostat-OECD/Hill: linking via the PPPs, i.e. maintaining fixity of the inter-regional PPPs. Comes down to calculating the linking factors as unweighted geometric means of the PPPs derived in the global calculation. Can be shown to minimize the changes to be made to the global PPPs to maintain fixity.

The choice between methods 1 and 2 is thus mainly a choice of priority of prices or volumes.

Again, Sergey criticized either method because the regions are not treated symmetrically. He proposed to calculate first average regional BH PPPs and weights and then to do the EKS on that to calculate the inter-regional PPPs. The arguments pro and con are the same as those at the BH level.