

Estimating national annual average prices in the ICP-Africa context: Back to surveys

Mathieu B. Djayeola & Michel Mouyelo-Katoula¹

Summary:

In most African countries participating in the 2005 ICP round, the collection of prices of goods and services consumed by households covered only the last 6 months of the reference year 2005. But national coverage and data collection during the entire period of the reference year 2005 are necessary to comply with the PPPs calculation requirements. If these conditions are not met then some adjustments are needed to come up with sound estimates of annual national average prices.

This article proposes a survey data validation approach towards the estimates of 2005 annual averages for the ICP-Africa participating countries. It suggests that intra-country data validation activities are conducted to a final stage in the participating countries. It concentrates exclusively on price survey and so frees survey averages prices of any inflation-type time dimension in spatial analysis context. It ensures the integrity of survey data for countries that undertook price collection throughout the year 2005.

Key words: *International comparison - price survey - data validation*

Résumé:

Dans la plupart des pays africains participant au cycle 2005 du PCI, la collecte des prix des biens et services consommés par les ménages s'est déroulée uniquement sur les 6 derniers mois de l'année de référence 2005. Mais la couverture nationale et la collecte de données durant la période entière de l'année de référence 2005 sont des conditions nécessaires pour se conformer aux exigences de calcul des PPA. Si ces conditions ne sont pas réunies, alors quelques ajustements sont nécessaires pour obtenir des estimations des prix moyens nationaux annuels cohérentes.

Cet article propose une approche de validation de données d'enquêtes conduisant à l'estimation de moyennes relatives à l'année 2005 pour les pays participants au PCI-Afrique. Elle suggère que des activités de validation de données intra-pays soient menées à une étape finale dans les pays participants. Elle porte exclusivement sur l'enquête prix et évite ainsi l'interférence de l'inflation résultant de la dimension temporelle dans l'analyse spatiale sur les prix moyens. Elle privilégie les données d'enquêtes pour les pays qui ont effectué la collecte tout au long de l'année 2005.

¹ICP-Africa is coordinated by Mr. Mouyelo-Katoula Michel. Mr. Djayeola B. Mathieu played a key role in the implementation of data quality initiatives.

Mots clés: *Comparaison internationale - enquête prix - validation de données*

1. Background

The International Comparison Program (ICP) is a global statistical initiative established in 1970 to produce internationally comparable price and expenditure levels to facilitate cross-country comparisons of GDP and its sub-aggregates in real terms and free of price and exchange rate distortions. From an initial number of ten countries that participated in the program in 1970, the ICP has grown into a worldwide program with some 140 countries participating worldwide.

The African Region launched ICP field operations with three countries namely Egypt, Mauritius and Zimbabwe starting collecting commodity prices in January 2005. Some countries began their field operation in May the same year while the remainder started in June-July.

Box 1 Quaranta Tables

“The Quaranta tables are quality control tools used (...) to identify and correct outliers in price data provided by countries participating in the Program. Named after their creator Vincenzo Quaranta of National Institute of Statistics, Italy, the Quaranta tables are designed to provide an indication of consistency with respect to basic headings, and to provide an analysis of the variability of surveyed prices, between items and across countries. Both the Eurostat and the OECD use the Quaranta tables in their PPP program”.

Source: World Bank: ICP News Vol 2, No. 2, April 2005

Earlier in June 2005, acknowledging the importance of an intra-country price data validation tool, the AfDB based ICP-Africa regional office developed the first version of the Semper² Validation Software and provided it to the countries to crosscheck their field data before onward submission.

²The Semper Validation Software is a user-friendly application to assist countries in their price survey data validation activities.

In December 2005, building on more than six month of data collection, the need for monthly time series analysis was emphasized by a number of participating countries. In response, the African Development Bank appended the Semper with a Time Line procedure. Using canned trends analysis – far from any inflation analysis – the Time Line set a framework for spatial and time data analysis thus providing a time dimension to Quaranta based analysis.

This paper proposes a survey data validation approach towards the estimates of 2005 annual averages for the participating countries. It ensures the integrity of survey data for countries that undertook price collection throughout the year 2005.

2. Objectives

The analysis of price spatial indices in the context of internationally heterogeneous consumer price indices points to the need to address critical issues related to price variability and consistency over time and space. The technological context in which the ICP-Africa was conducted required the development of ad-hoc applications comprising specific data capture, structure, and statistical validation modules. A few technical constraints needed to be addressed with respect to the amount of data to be handled.

3. Analyzing spatial data with inflation-type indicators

According to the literature and empirical ICP work, the analysis of ICP data with an inflation factor usually shows semantic problems related to the nature of ICP indices being spatial and international in opposition to the use of temporal indices for inflation measurement in a purely national economic context³. With experience, other approaches were proposed in terms of considering ICP prices in their original essence, considering that even if the prices were collected on a monthly basis, they are meant to be used for spatial analysis.

The 2005-2007 round of ICP data collection has demonstrated that when the product list is long and involves a group of countries with heterogeneous economic situations⁴, price variation is the rule and not the exception. This calls for a sound methodology of price averaging.

3C.f. for example Alan Heston and Robert E. Lipsey in International and Interarea Comparison of Income, Output, and Prices, National Bureau of Economic Research at the University of Chicago, Chicago and London, 1999.

4The survey covered 853 products clustered in 113 basic headings and prices were collected in a continuum of such different economic and cultural environment like in Tunisia, South Africa, Senegal and Ethiopia, and forty countries between.

The proposed data validation process includes a number of steps that need to be conducted with care as the current round of data collection was made mainly processed under Microsoft Excel environment which is limited in terms of number of records that can be processed simultaneously.

4. Major technological constraints

Intra-country data validation in the African context is based on one monthly spreadsheet containing all records from field surveys. There is a limitation of about 65,000 records that can be processed on a single spreadsheet under Excel 2003 and earlier versions. Available computers also showed some limitation in terms of memory when the number of records to be processed using the Semper approach is more than 30,000. These technical constraints suggest that data collected by the countries should be reorganized for effective processing.

All country data collected in 2005 and 2006 have to be combined into a single database and processed with the same means as when each country was validating data on a monthly basis. Concerning the scope of data to be analyzed, the minimum number of monthly records for this round is about 3,000 and the maximum is above 40,000 with an average around 11,000. All in all, five million records are included in the database.

For quality purposes, the treatment of such a large database requires a step by step approach to strictly comply with database integrity and data preservation at each stage.

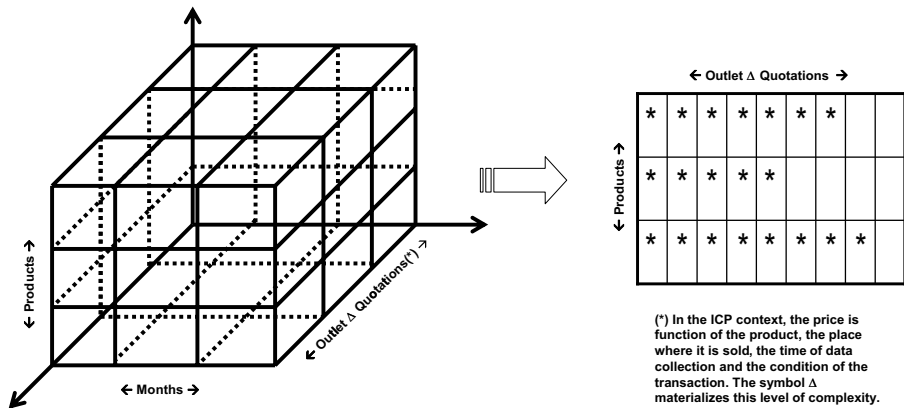
Experience from selected countries ranging from the lowest to the highest in terms of price quotations suggests that it takes a day to manage the resulting database. As concerns the duration of the data validation process, it depends on the speed of carrying out monthly time series, and quarterly data analysis. All in all, one week of work would be enough at the country level to generate a workable price database.

5. The proposed approach

The proposed approach consists of five steps: (1) describing the methodology for reorganizing the data validation process, (2) determining the storage platform, (3) defining the best procedures for data integrity and trace-ability, (4) undertaking the data validation process, and (5) setting a closure rule to ensure that no data was lost during the data handling process. These steps are described in the following sections.

5.1. The methodology

The original configuration of price data collected by countries followed a cube presentation with rows representing the products, columns the months and depth for price quotations. Each month (Dimension 1), price quotations from surveyed outlets (Dimension 2) were analyzed against the corresponding information from other months, as far as a particular product (Dimension 3) is concerned. A monthly average price was then calculated per product: the price is in a cubic graphic represented by these three dimensions.



Graph 1: ICP-Africa survey databases –From cubic to matrix representation

In the proposed structure, all the twelve monthly databases are merged into one annual database. For each product (Dimension1), all outlet price quotations (Dimension 2) for the whole year are checked against one another, and annual average can be easily computed.

The approach is holistic in the sense that it deals with the entire product list, with all months and all outlets. It ensures data integrity while complying with statistical principles such as spatial and time representativity, and also, it is free of non survey inferences.

5.2. A holistic platform

The proposed platform is based on the extension of the CountryData concept⁵ used by the countries during the field operations. The main idea

⁵The CountryData file provides the standardized structure for price collection database at the country level. It contains information on the product description, the quantity, the price, the outlet, and the price collection center identification. It serves as input to the Semper Validation Software.

is to create a super CountryData containing all price quotations from a particular country. If for logistical reasons (memory, disk space, and file capacity etc.), more than one file is needed, a partitioned CountryData can be created.

Ideally, to launch the process, a twelve-spreadsheet CountryData file must be created for each country. Each sheet refers to one month. Of course, countries with more than twelve months of data collection would need a larger number of sheets.

This approach can be extended to a regional process aimed at creating a regional database.

5.3. Data integrity and trace-ability

The records coding structure used by the countries is a sequential series of five digits index ranging from 00002 to merely 50000 in a country with a very high number of quotations. It is suggested that for each month, the code should be processed by adding a four-digit prefix. The first digit (5 for 2005 and 6 for 2006) identifies the year of data collection. The code also includes a month identifier in the form of a three-digit sub-code such as APR for April. Whatever language is used for data collection, the following are the monthly identifier to be used: JAN - FEB - MAR - APR - MAY - JUN - JUL - AUG - SEP - OCT - NOV - DEC.

The last part in the string is the original five digits index in the Country-Data file. The resulting nine-position string ranges from 5JAN00002 to 6JUN50000 for each country.

5.4. Data validation

The database can be safely sorted by product categories. A back-up of the CountryData file is critical at this very stage. The reason is that here start non reversible data management processes. Sorting procedures are usually based on the fastest process to reorder records in databases and the subsequent algorithms do not provide room for reverting to the original database in case of need. It is the responsibility of the data manager to ensure the reversibility. For this, one of the most common options is a good indexing system such as the one described in stage 3 above.

Once the database is sorted, it is proposed to split it in as many portions as necessary for onward validation. Each portion can relate to bunches

of ten basic headings: 001 – 010, 011 – 020, 021 – 030, ... 090 – 113. In the hypothesis of twelve months of data collection, depending on the country specifics, the new CountryData file will have between one and ten spreadsheets.

At this point, it is important to highlight a feature of the Semper Validation Software not yet fully applied by the countries. The software treats only the active sheet of the CountryData file, this means that if price quotations cover several sheets, the user can implement the validation on each individual spreadsheet.

This process ensures that the resulting CountryData file is clean of any unnecessary aspect and brings back to the spatial objective of ICP. Once this is secured, one can split the database in as many parts as required. For instance, 2005 data can be separated from 2006.

It is also suggested to set a threshold for price variation in the database. The mostly used indicators for variation analysis in the Semper context are the Outlier indicator⁶ in the Semper output file (ValidationBook) and the coefficient of variation which scales the Standard Deviation to a percentage of a product average price.

It is further suggested that no ValidationBook should contain any Outlier indicator beyond 500 in absolute terms. This is in line with ICP practice and leads to a reasonable level of coefficient of variation. In particular, it avoids cases where commas are misplaced or misprint such as 1500 instead of 150 or 15000 – these are known as “zero” typing mistakes.

5.5. Closing mechanism

The user may need to verify that no data was lost during the database reorganization process. This is important for data integrity and quality purposes as database handling is a risky process unless it is surrounded by tight verification procedures. Selected indicators can be used in this regard. Two of them are listed below:

⁶The Outlier Indicator in the Semper Validation software is a mix of central tendencies and dispersion indicators combining the Mean, the Minimum, the Maximum, and the Standard Deviation. It produces an index number as high as a price diverges from the average of the product and is scaled by tens to alert on the observed arithmetic distance.

- (1) the number of records before and after running this process, and
- (2) the price levels obtained in national currency at the beginning and at the end of the process.

Estimated average prices need to be individually checked against the known price level for each product on the item list.

In addition to this, a list of tasks aiming at completing the validation process at the national level is proposed in the following section.

5.6. A quick checklist of proposed tasks

The checklist includes the following tasks:

- a. Have a single Excel file containing all monthly country data - each sheet containing one month of data - and re-code the records accordingly. Keep a clean copy of this file.
- b. Sort the database by product code to bring all price quotations for a particular product together.
- c. Create a dedicated database for product groups – a cluster will include 5 to 10 basic headings, but not more than 30,000 records per sheet.
- d. Check the integrity of the resulting file by ensuring that the number of records is the same as before the beginning of the process. Make sure that a copy of the initial file was secured before starting the data validation process.
- e. Validate the data in the resulting country data file following the same rule as was done during data collection activities.

Conclusion

How suitable is it to use six months data collection for annual averaging purposes? This question certainly goes beyond the time dimension and, apart from the fact that in 2005, prices were mainly surveyed in the second half of the year, we could express it the other way round: How suitable is it to use survey data to estimate actual unknown values?

In the past ICP rounds and also in other regions for the current ICP round, survey activities were carried out only during few months in selected areas (quarterly, one survey over three months, only in capital cities, in major cities, etc.). The robustness of methodology and experience showed that under controlled conditions, this sampling approach does not affect the quality of the PPPs.

In the African ICP exercise, all participating countries have survey data for at least twelve months but these months were spread over two calendar years. It is also established that all participating countries have CPI data for basic categories for which ICP results will be published.

To put more emphasis on survey data, a simple back-cast methodology for missing months in 2005 is to adjust the corresponding 2006 survey data with 2006/2005 CPI indices. Each monthly database would therefore be adjusted accordingly.

References

African Development Bank. (2005). *Training of Field Supervisors and Price Data Collectors, A Trainer's Guide*, ADB, Tunis, Tunisia, 1-77

Alan H. and Robert E. L. (1999). *International and Interarea Comparison of Income, Output, and Prices*, National Bureau of Economic Research, University of Chicago, Chicago and London, 1 - 530.

Astin, J. (2004). *ICP 2004 Operational Manual, What National Coordinators Need To Know*, Canterbury, UK, 48 - 50.

Djayeola B. M. & Rittenau R. (2006). Experiencing *the Semper Validation Software: Genuine African solutions for data validation within the International Comparison Program*. The African Statistical Journal, Tunis, Tunisia, 32 p.

Eurostat-OECD (2005). *PPP Methodological manual - Annex IV Quaranta editing procedure*, Luxembourg, Luxembourg, 1-6

Rittenau, R. (2005). *The quality of data in ICP-Africa*, ADB, Tunis, Tunisia, 1-6

Sergeev, S. (2003). *Description of the VBA program for the computation of the EKS-PPP at the basic heading level and the "Quaranta" tables*, Vienna, Austria, 1 - 27.

World Bank (2005). *ICP News: An e-newsletter for the International Comparison Program*, Volume 2, No. 2 - April 2005, Washington, USA, 1 - 5.