**Assessing the uncertainties of statistical outputs: the case of purchasing power parities**

Laurent Olislager, Eurostat, Laurent.Olislager@ec.europa.eu

**Abstract**

*Purchasing power parities (PPPs) are indicators of price level differences across countries. They are used for a wide range of purposes. In the Eurostat-OECD PPP exercise, prices are collected via specifically-designed surveys, for a variety of products and services covering an important subset of economic activities. For each country and item, the average price of the item is compared to the prices of comparable items in other countries. PPPs are obtained by averaging all the bilateral price ratios for products and services grouped in pre-defined categories. Using weights from National Accounts, PPPs are aggregated up to the GDP level.*

*The methodologies of the surveys are designed to guarantee comparability and reliability of the results, which are subject to a thorough validation process incorporating various quality measurements. However, due to the use of purposive sampling and the multilateral nature of PPPs, it is not possible to calculate precise error margins for PPPs. Experience suggests that differences between countries of over two percentage points are generally statistically significant. In this work, we aim at better characterizing the uncertainties of PPPs. We have carried out various sensitivity analyses to assess the robustness of the results when subject to changes in the input data and/or implicit assumptions in the calculation. General trends confirm the high quality of the results and can contribute to an enhanced communication with stakeholders. Our results also allow to assess the relative impacts of the input data on the final results; it can contribute to enhancing the efficiency and effectiveness of the production process, by helping in the optimization of the design and validation.*

**Keywords:** international comparisons, price statistics, purchasing power parities, uncertainty estimation

**1. Purchasing power parities**

Purchasing power parities (PPPs) are indicators of price level differences across countries. They are used for a wide range of purposes, such as direct comparisons of price level indices (PLIs) and their dynamics among countries, and inter-country comparisons of gross domestic product (GDP) and its component expenditures in real terms – i.e. adjusted for the geographically heterogeneous price levels of goods and services. PPPs are by essence a multilateral construct and interpreting their meaning must be done cautiously, especially when considering series over time (in particular since their geometric mean over countries is set to a fixed value).

PPPs are calculated for the 37 countries involved in the Eurostat-OECD PPP exercise. Prices are collected via specifically-designed surveys, for a variety of products and services covering an important subset of all economic activities. For each country and item, the average price of the item in the considered country is compared to the prices of comparable items in each of the other countries. The PPPs are obtained by averaging all the bilateral price ratios for products and services grouped in pre-defined categories – in accordance with the final expenditure classification of the European Standard of Accounts (ESA 2010). Using expenditure weights from National Accounts, PPPs are aggregated to analytical categories, up to the GDP level, and disseminated in Eurostat database.

PPPs are calculated using the Elteto-Koves-Szulc (EKS) method, consisting of a number of steps (calculation of Laspeyres-, Paasche-, Fisher-type PPPs, "bridging" of countries, procedure to obtain transitivity) that are described in chapter 12 of the Eurostat-OECD Methodological Manual on Purchasing Power Parities (Eurostat & OECD, 2012), hereafter "the PPP manual".

**2. Precision and accuracy of purchasing power parities**

As regards precision and accuracy, the PPP manual, in section 1.4.2, states that "*it is not possible to calculate precise error margins for PPPs or for the real expenditure levels and price levels derived from them*". "*Experience suggests that differences between countries in these indices of over two percentage points are generally statistically significant*", at GDP level – for lower levels of aggregation, expected error margins are wider.

Evidently, the precision and accuracy of the output data – the PPPs and the real expenditure levels and price levels derived from them – depend on the precision and accuracy of the input data – surveyed prices and expenditure weights, but also assumptions that are, to some extent, arbitrary, e.g. inclusion of selected goods and services in the purposive sampling and assessment of their representativity.

Our purpose is twofold. First, we aim at better characterizing the output data, to refine the analysis of the results and their trends, and enhance the communication towards stakeholders. Second, we aim at better assessing the relative impacts of the inputs on the outputs, to help optimizing the statistical production process, particularly in the design and validation phases of the surveys – as described in the next section.

**3. Relative impacts of the input data**

*3.1. Exclusion of items*

Deriving PPPs starts with the collection of prices for goods and services comparable in several countries. In the vast majority of cases, whether an item belongs to a category or another is uncontroversial; on the other hand, including in the sample an item rather than another is a compromise ultimately decided by expert judgment, and surveying or not an additional item is a trade-off between cost and expected benefit to the results. To estimate in a simple way the marginal information gain of an item, we calculate the deviation of its category's results when excluding it from the input data for all countries. The absolute average impact is found to be 1.2% for consumer goods and services, considered hereafter. Detecting items with a huge impact – deleting some may actually prevent results to be calculated by lack of data – should lead, in design or validation, to a renewed focus on their category, e.g. by surveying more items. Indeed, iteration through all items shows that, as expected, the marginal impact decreases for an increasing number of items in the category, by 0.015% per item on average – outliers may trigger an in-depth analysis. The impact can be linked to other parameters, e.g. dispersion of results, through various regressions.

*3.2. Representativity of items*

While the decision to include, or not, an item in the survey sample is multilateral, each country must separately assess the importance of surveyed items. However, with the exception of products covered by transaction data, expenditure weights are usually not available at item level. The calculation of PPPs at the lowest level of aggregation is thus "quasi-weighted": items nationally marked as "representative" carry a higher implicit weight. Representativity assignment can be rather arbitrary, since the very reason to use it is the lack of data on sales at item level. To assess its impact, we compare the PPPs with an unweighted calculation, i.e. considering all items representative. Big deviations indicate heterogeneous assignments among countries that should reflect national differences in consumption patterns. It may lead to comparability issues to be analysed. Iterating through items allows estimating the absolute average impact of representativity assignments on the PPPs (2.8%), which decreases for an increasing number of items per category.

*3.3. Fluctuation of prices*

National average prices used to calculate PPPs are derived from collected prices. The sampling is informed by various sources but, since market shares are usually unavailable at item level, expert judgment is needed. Bias in the sampling would mechanically lead to bias of average prices and bias of PPPs; e.g., if a country over-estimates its prices by x%, the ratio of its PPP (over-estimated) to others (slightly under-estimated in compensation) will be over-estimated by x%. A major purpose of the extensive validation process of the PPP programme is to avoid such cases and guarantee the accuracy of the results; this can only be done on a case-by-case basis.

Because samples are not representative, no direct calculation of the precision on PPPs can be performed, but we can numerically introduce uncertainty on the prices and estimate the impact on the PPPs, with a Monte-Carlo simulation. For each run of the experiment, each average price is randomly selected:

$$modified price=reported price×\left(1+rand\right).$$

The interval around the reported value is tuned by the random parameter, whose distribution is symmetric around zero and whose maximal value can be fixed or made dependent on the item and country, according to parameters such as the number and dispersion of prices collected. Each run gives rise to a set of PPPs deviating from the reference case; after a sufficient number of runs (typically a few hundred), the results converge on average and we can evaluate their dispersion. For example, the normalized average standard deviation of the PPPs at the lowest level of aggregation is approximately 3% if prices fluctuate in a "flat" interval of ±10% around their reported value, demonstrating the relative robustness of the results. Again, the deviations decrease with increasing number of items, and outliers may trigger an in-depth analysis in the validation and design phases of the survey.

*3.4. Exclusion of countries*

The PPP of a country is obtained by averaging bilateral price ratios and thus depends on the input data from all other countries. Withdrawing a country leads to a global re-scaling of the PPPs of the other countries and to modifications of individual PPPs and their ratios. Importantly, those effects take place when the list of countries constituting the base of the comparison – the European Union – is modified.

**4. Characterization of the output data**

Evaluating the precision and accuracy of PPPs through evaluation of the precision and accuracy of the input data and propagation of errors is not rigorously possible. We argue that something sensible can nevertheless be said on the precision and accuracy of PPPs, and that it is of value for both producers and users of PPPs.

We can derive error margins for PPPs at all levels of aggregation, up to GDP level, from error margins on the input data, namely prices and expenditure weights – using a Monte-Carlo simulation, similarly to section 3.3. Numerous limitations bound the reliability of the results, but they can be made "good enough".

First, error margins on average prices are unknown. However in some cases they can be estimated indirectly via comparison with a more complete source, e.g. newly available transaction data. Those cases then provide an empirical link between error margins and parameters such as the number and dispersion of prices collected.

Second, error margins on expenditure weights are also unknown – we expect them to get smaller with bigger levels of aggregation. However active research in the field of national accounts has started to provide reliable estimates. Also, the PPPs are less sensitive to fluctuations of weights than fluctuations of prices, so that a conservative estimate can be used without too much impact on the total error.

Third, some categories of GDP are not covered by a price survey of the type used for consumer goods and services. Estimating error margins for those PPPs, calculated or imputed using other sources, is thus usually even more challenging. The error margins themselves have to be calculated or imputed from error margins more readily computable. The same applies for gaps due to data unavailability. This procedure may not provide a sufficiently reliable estimate of the error for those categories, but allows deriving an estimate for aggregate categories.

Fourth, PPPs are multilateral constructs with many possible mathematical definitions and properties. Propagating errors within a given methodology – the EKS method – does not provide a conservative estimate of the error margin for PPPs. Calculating PPPs with other methods therefore provides a needed "sanity check" of the results. Challenging assumptions (see e.g. sections 3.1 and 3.2) serves the same purpose.

All in all, we deem it possible to provide error margins for PPPs incorporating most of the uncertainty, though they need to be defined and used with care.

**5. Conclusion and perspectives**

We have shown that simple numerical experiments can inform the producers of PPPs how the output data are impacted by changes to input data and related assumptions. Our approach can be used as complement to traditional indicators used in the design and validation phases of the PPP surveys, by helping very concrete decisions such as re-balancing item lists from a category to another or re-focusing validation efforts from an outlier to another, with the goal to improve the efficiency and effectiveness of the process. Our approach can also be used to derive error margins for PPPs and the real expenditure levels and price levels derived from them. In this regards, our results are only preliminary and ad-hoc assumptions need to be refined. Such an approach comes with intrinsic limitations, and expert judgment will always remain necessary to conduct the PPP process and to analyse the results, mainly in terms of plausibility. However, "good enough" estimations of error margins, if well communicated to users, may ultimately reinforce the confidence in the statistical product.

**6. References**

Eurostat & OECD (2012), Eurostat-OECD Methodological Manual on Purchasing Power Parities. Available at: <http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-12-023> (Accessed: 20 May 2018).

**7. Acknowledgements**

The author expresses his gratitude to Paul Konijn and Marjanca Gasic, Eurostat, for their continuous support and for their valuable comments and suggestions.