

Unclassified

STD/TBS/WPTGS(2011)3

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

27-Oct-2011

English - Or. English

STATISTICS DIRECTORATE

Working Party on International Trade in Goods and Trade in Services Statistics

MEASUREMENT ERRORS DETECTION IN THE ITALIAN EXTERNAL TRADE UVIs

7-9 November 2011, OECD Headquarters, Paris

This paper is for discussion. It has been prepared by Paola Anitori for agenda item 5.1 "UVI calculations in Italy".

Contact person: Paola ANITORI, E-mail: anitori@istat.it

JT03310038

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format



STD/TBS/WPTGS(2011)3
Unclassified

English - Or. English

MEASUREMENT ERRORS DETECTION IN THE ITALIAN EXTERNAL TRADE UVIS

Paola Anitori – ISTAT

Premise.

In 2006 a new method for outlier detection and correction was implemented in the Italian external trade UVIs in order to deal with measurement errors and outliers.

The method was applied to the distribution of the elementary unit values (levels) in each stratum defined by product-flow-country of origin/destination with the aim to control for the variability inside the strata, mainly due to errors on reported quantities. It is based on a non-parametric algorithm for univariate, skewed distributions, *ad hoc* adapted to take into account some peculiarities of the variables involved.

This paper explains the method and discuss the improvements achieved by the new UVIs series in terms of reliability and accuracy.

Background information on UVIs in Italy.

1. External trade UVIs in Italy are chained Fisher-type indexes disseminated according to the breakdown shown in table 1.

Table 1. UVIs dissemination breakdown in Italy.

Product	Area of origin/destination
• CPA 2-digit	WORLD, EU, EURO AREA EXTRA-EU
• CPA 3-digit	
• MIGs	
• Total	World, EU, Euro area, Extra-EU,

2. Each monthly link is calculated as the squared root of a Laspeyres-type and a Paasche-type index both based on the previous year. On their turn, each Laspeyres and Paasche link is a weighted average of elementary unit value indexes in each stratum given by “product-country of origin/destination-flow”, the products being classified according to the Combined Nomenclature at 8-digit level. Fisher links are then chained back to a reference year (currently 2005=100). The total number of elementary strata is around 220.000.

3. The Fisher type index allows to derive quantity indexes as the ratio between a Fisher value index and the Fisher “price” (UVI) index according to the well known property $IV=IP*IQ$ the IQ being the Fisher quantity index.. Istat currently disseminates monthly quantity indexes according to that formula.

The Italian outlier detection and correction method.

4. As it is well known, a Unit value Index is an estimate of the rate of change over time in the “price” of exported and imported commodities. In a strict statistical sense¹ UVIs are not “pure” price indexes as they are potentially biased (with respect to prices) due to both changes in time in the mix of items reported in custom documents but impossible to detect, and the poor quality of the recorded quantity information related to each transaction.

5. Nevertheless, UVIs are used in several countries as the only price measure in international trade or as complementary estimates to proper export/import price indexes² (XMPI) because they are a ready available and low-cost by-product of trade data.

6. UVIs compilers know how important is to have a statistical tool able to reduce their potential bias in order to get more reliable estimates. Although it is impossible to remove important drawbacks such as items specification and identification, it is possible to make significant improvements by means of outlier detection routines that enable to isolate potential errors and abnormal values, which can be subsequently corrected or discarded according to a chosen strategy.

7. In Italy the NSI decided to introduce a more sophisticated outlier detection method when a new EU Regulation - requiring quantity data expressed in supplementary units - came into force in 2006 and compelled to move from an HS 6-digit level based methodology to a CN 8-digit elementary indexes. At this level of detail, measurement errors relating to quantities became more evident marring the reliability of the index at the very first stage of calculation by producing a large amount of abnormal values.

8. Due to the impossibility of referring back to the firms for all the suspected outliers, it was decided to implement a statistical deletion routine chosen among those based on non-parametric methods for univariate, skewed distributions. Most of the solutions within this class of methods, in fact, take into account the asymmetry of the distribution of the variable and, using simple robust parameters like quartiles, median etc., they set the thresholds above which an observation is regarded as a measurement error according to the shape of the distribution itself.

9. The method is an extended version of the well-known Asymmetric Fence Method (AFM), integrated *ad hoc* in order to take into account some peculiarities of the variable. The whole set of transactions (that is each movement registered by Customs) was stratified by product (CN 8-digit), country and flow; then the log-distribution of the elementary unit values (levels) referred to each transaction and recorded monthly in each single stratum was considered. Given the evident asymmetry of the distributions, it was advisable to look for an appropriate transformation of the raw data by means of the generic Box-Cox algorithm and after several experiments, the most appropriate transformation both for import and export data resulted to be the log-transformation.

10. The method was applied to the unweighted distribution of the elementary unit values in each stratum according to the hypothesis that measurement errors happen independently from the relative “importance” of the observation in the stratum.

11. As far as the AFM is concerned, according to its general formulation, given a set of observation $\{x_1, \dots, x_n\}$ ³, the method regards the single x_i as an error (or outlier) if one of the following conditions is verified:

$$q_1 - x_i > k_{AFM} * \max(q_2 - q_1, c * |q_2|) \quad (2)$$

¹ See IMF, 2009 for a complete treatment of the issue.

² A proper XMPI is a survey based index.

³ Here x_i is the logarithm of the elementary unit value.

or

$$x_i - q_3 > k_{AFM} * \max(q_3 - q_2, c * |q_2|) \quad (3)$$

with q_1, q_2 and q_3 respectively the first quartile, the median and the third quartile of the distribution of x and k_{AFM} and c are two parameters to be determined (ILO, 2004): in particular, the former defines the width of the interval outside which outliers are identified and the latter avoids problems which can arise when the distance between the quartile and the median is too small (*e.g.* if the variable has a very low variability and the distribution is highly leptokurtic). According to our experience, the optimal choice for our purpose was $k_{AFM} = 1$ and $c = 0.05$. This method has a breakdown point (that is the percentage of outliers that can be detected without altering the estimates) of 25% (Geyer, 2006)⁴, and since the interval is larger on the side where the distance between the quartile and the median is larger, the AFM respects the natural asymmetry of the distribution.

12. Nevertheless, the method is not efficient if the number of observations in the stratum is too small,⁵ this number being recognised to depend on the phenomenon analysed, even if some literature fixes it at $n=50$.⁶ We observed that, as far as unit values are concerned, the number of observations in a stratum is quite random, not depending on systematic factors. After some empirical tests we decided to set $n > 100$ per stratum. If $n \leq 100$ we applied a different non-parametric method based on the Mean Absolute Deviation (MAD) :

$$MAD_i = \underset{i \in nc}{\text{median}}(|x_i - q_2|) \quad (4)$$

The single transaction x_i is considered as an extreme value if the following conditions are both satisfied:

$$\begin{cases} |x_i - q_2| > k_{MAD} * MAD \\ \frac{|\exp(x_i) - \exp(q_2)|}{|\exp(q_2)|} > A \end{cases} \quad k_{MAD} = 2 \text{ e } A = 0.1 \quad (5)$$

where A and k_{MAD} are fixed by empirical tests. This method is more robust than the AFM (its breakdown point is 50%) and, in our case, it has proved to be not excessively severe if the distribution is unimodal. Moreover $k_{MAD} = 2$ is coherent with $k_{AFM} = 1$ because of the approximated relation between the interval $\text{median} \pm 2 * MAD$ in the asymmetric distributions and the interval $\text{mean} \pm 1.3\sigma$ in the Normal distribution⁷, the latter being the confidence interval of a standardised Normal including 80 per cent of accepted values.

13. Then all transactions in a month m of year t or in the base year $t-1$ failing the AFM test ($n > 100$) or the MAD test ($n \leq 100$) were flagged and excluded from the calculation of the elementary index of the stratum. In other words, the elementary index

- 4 If $x \rightarrow N(0, \sigma)$, quartiles and median satisfy the following property: $q_3 - q_2 = q_2 - q_1 = \frac{q_3 + q_1}{2}$. Therefore, if $c = 0$, the AFM method reduces to the most known Tukey method. When $k_{AFM} = 0$, exactly 50% of the observations fall outside the interval. Any variation of k_{AFM} implies a variation in the number of detected outliers.
- 5 If $n=5$, a single outlier would affect the quartiles' position making the method useless.
- 6 In our experience differences between $n= 50$ and $n= 100$ proved to be negligible: so, we decided to set $n=100$ as a precaution.
- 7 In the Normal distribution the approximated relation $\sigma \cong 1.5MAD$ holds and the mean coincides with the median.

$${}_{t-1}I_m^{nc8,t}$$

referred to month m , year t and base year $t-1$ in a single stratum (here for simplicity suffixes referred to country and flow are omitted) is given by the ratio of non-flagged monthly unit values ${}^{*nc8,t}vmu_m$ and non-flagged base year unit values ${}^{*nc8}vmu_{t-1}$ in the same stratum:

$${}_{t-1}I_m^{nc8,t} = \frac{{}^{*nc8,t}vmu_m}{{}^{*nc8}vmu_{t-1}} = \frac{\sum_{i \in nc8} val_{i,m}^{*nc8,t}}{\sum_{i \in nc8} q_{i,m}^{*nc8,t}} / \frac{\sum_{i \in nc8} \sum_m val_{i,m}^{*t-1}}{\sum_{i \in nc8} \sum_m q_{i,m}^{*t-1}} \quad (6)$$

14. The exclusion of flagged transactions from the calculation of the CN elementary indexes corresponds to giving these transactions a null weight. No consideration on the weight (in value) of the potential outlier was taken into account at this level: we judged that measurement errors can affect the variable regardless to the importance of the transactions within the stratum. Random checks made on the set of the rejected records proved that in the majority of cases non-congruence due to errors of different kind was confirmed even at firm level. In our opinion these exclusions cannot be regarded as a total loss of information, in as much as the weight of the transactions excluded from the UV elementary indexes will enter the calculation of the aggregated index numbers (e.g. at CPA or SITC levels) of values and volume⁸ computed by means of the Fisher formula.

15. At this stage, the methods returned a percentage weight of flagged transactions, at CPA 2-digit level, between 1% and 25% depending on the products: raw materials or intermediate goods, usually homogeneous goods, showed in fact lower percentages of outliers detected.

16. We applied the same routine to the elementary unit values used in the base year of each link (denominator of the index) in order to maintain the coherence of the ratios.

17. The deletion choice was somehow constrained by the way the monthly production flow is organised in order to produce the press releases. At this stage checks are made only on very big values and quantities at aggregated levels and there is not time for the staff to check with firms the amount of all suspected values detected by the routines.

18. At the same time no imputation method is applied to replace the deleted observation in order not to introduce imputation bias in the resulting distributions. Moreover, due to the high number of strata we judged not economically wise to implement an imputation routine. As we previously specified, delete unit values are taken into account by means of maintaining their original weights in the calculation of the aggregated indexes.

19. Aggregated results of the exercise are shown in table 2 where the CV outlines a reduction in total variability of the unit values distribution by over 50 per cent.

8 The calculation of the aggregated indexes of value, in fact, is based on the "full" values of each stratum (i.e. the monthly figures comprise the values of all transactions, the flagged ones included). The correspondent aggregated index of volume - given by the ratio between the index of value and the UVI - will then reflect the full range of events recorded in the stratum.

Table 2 Evaluation of the effect of Outlier detection method on unit value distribution by flow. Total trade. Years 2005-2007 (average annual values)

YEARS	outliers (% number)	Total loss (1) (% value)	% reductio n in the CV (2)	INITIAL DISTRIBUTION		ENDING DISTRIBUZION (NET OF ERRORS)	
				Asymmetry	IQR std range (3)	Asymmetry	IQR std range (3)
				Imports			
2005	20.7	18.2	52.5	1.5	1.9	0.5	0.5
2006	20.6	18.1	53.2	1.5	2.0	0.5	0.6
2007	20.9	18.1	55.5	1.5	4.3	0.5	0.5
				Exports			
2005	20.6	20.6	50.6	1.5	3.1	0.5	0.6
2006	20.6	20.1	50.8	1.5	3.4	0.5	0.6
2007	20.5	20.7	52.5	1.5	2.9	0.5	0.5

Source: ISTAT – Foreign Trade Statistics

(1) Percentage value corresponding to the discarded unit values

(2) After discarding errors

(3) Standardised range given by the ratio between inter-quartile distance and the median

20. Moreover the *ex post* distributions appear corrected and symmetrised and this condition makes results statistically more “robust”.

Conclusions.

21. Custom data are probably one of the best examples of a cost-effective way of producing information through administrative records. This explains why UVIs, as a simple by-product of values and quantities reported on custom documents, are still calculated by a lot of countries despite their many drawbacks when used as substitutes of export and import prices. Nonetheless, Statistical Offices have to do their best to reduce potential bias and to ensure their reliability and accuracy.

22. Measurement errors and outliers management is certainly a way of reducing the variability traditionally affecting UVIs. In particular, reported quantities are traditionally affected by errors that mar the calculation of unit values at very elementary stages.

23. From 2006 ISTAT implemented a statistical tool to tackle this problem. By adapting the AFM method for univariate and skewed distributions to the empirical data set, we managed to reduce the total variability in each stratum by 50 per cent, improving the reliability of the resulting index..

REFERENCES.

Anitori P, M.S. Causo (2008), *Outlier detection and treatment: quality improvements in the Italian unit value indexes*, Seminar on External Trade data production, Eurostat 12-13 December 2007.

IMF (2009), *Export and Import Price Index manual – Theory and practice*, Washington, Usa.

Thompson, K.J. (1999), *Ratio edit tolerance development using variations of exploratory data analysis (EDA) resistant fences methods*, 1999 FCSM Research Conference Papers.

Tukey J.W. (1977), *Exploratory data analysis*, Addison-Wesley, NY.

UNSD (1981), *Strategy for price and quantity measurement in external trade - A technical report*, Statistical papers n°69, series M, NY.