32nd Voorburg Group Meeting

New Delhi, India

23-27 October 2017

Cross-cutting topic

Quality Indicators: a proposal to assess quality for Services Producer Price Indices

Cristina Cecconi, Fabrizio Marinucci, Francesca Sansone

Istat, ITALY

Table of contents

Introduction[1]

In Italy, the development of a set of Quality Indicators for Short Term Statistics (STS) is regularly requested by Eurostat that systematically monitors and measures quality for Principal European Economic Indicators (PEEIs)[2] among which Services Producer Price Indices (SPPIs) are included.

This paper provides a simple proposal to outline quality for SPPIs focusing on the quality of the statistical output choosing an ad hoc set of key indicators.

Paragraph 1 provides a brief overview of data quality assessment, Paragraph 2 presents quality criteria and indicators and contains a pilot example applied to Services Producer Price Indices, Paragraph 3 concludes.

1 Data quality assessment

A global data quality assessment should be done according to three quality aspects, closely interrelated: (i) characteristics of the statistical production process; (ii) characteristics of the output (statistical product); (iii) perception of the statistical product quality from the users' perspective. According to this view, focusing only on one of the previous aspects couldn't be a sufficient solution but, undoubtedly, monitoring all three together may be very expensive. Even if some methods (self-assessments, audits, certification, etc.) can cover all the aspects, others refer only to one aspect, for example: key process variables monitor the production process; quality reports and quality indicators measure product quality; customers/user surveys analyse user perception and satisfaction[3].

Though in the European Statistical System (ESS) priority is given to the statistical product quality aspect, assessed according to the ESS quality components[4], it can be useful to explain the three aspects mentioned above:

(i) production process quality hasn't an ESS standard definition as for product quality. Principles are more general. Key process variables are the variables with the largest effect on product characteristics and they vary by product quality component and by type of process. Typical process variables are: resources and time used, response rates and burden, complaints, interviewer performance, as well as error rates (for example in interviewing, editing and coding, disclosure control and dissemination);

(ii) product quality components are defined by Eurostat to regularly monitor the quality of short term statistics produced by European countries. They refer to six quality dimensions: Relevance, Accuracy, Timeliness and Punctuality, Accessibility and Clarity, Comparability, Coherence. They serve both to provide users with important background information on STS data as well as to allow Eurostat to identify good practices and needs for harmonisation of the national STS data production;

---

[1] Attributions to the authors: to C. Cecconi paragraphs 1, 2.2, 3; to F. Marinucci paragraph 2.3; to F. Sansone paragraph 2.1. The views expressed in this paper are those of the authors only and do not necessarily represent the position of Istat.

[2] A list of 19 short term indicators relevant for conducting economic and monetary policies for the Euro zone (http://ec.europa.eu/eurostat/web/euro-indicators).

[3] Eurostat. 2007. *Handbook on Data Quality Assessment Methods and Tools*. Wiesbaden.

[4] Quality components are commonly also referred to as criteria or dimensions. These terms are normally used synonymously.

(iii) statistical product quality can be perceived differently by users and by National Statistical Institutes (NSIs) even if the quality components are the same. Some of them are difficult to assess for the users (for example accuracy), on the contrary, others are easier and users can easily formulate their needs (for example accessibility or timeliness). Users' surveys normally get information on the users' perception to be used for improving quality. Users are various and have different expectations and satisfactions degrees, so NSIs carry on several surveys: traditional users' surveys directed to known users; surveys of confidence directed to unknown users; target groups' specific surveys (for important stakeholders, web-users, journalists) and so on.

In order to assess data quality it is indispensable to have appropriate tools: quality reports that describe data characteristics by quality components and provide documentation on the quality features of statistical products seem to be the most suitable means. In fact quality reports are important both for producers of official statistics as well as for users. In fact, on one side, NSIs systematically need to have a picture on statistical product quality in order to understand the strengths and weaknesses of statistics and to identify the improvements they can do; on the other side, reports help users to use statistical outputs properly.

2. Quality dimensions and indicators.

2.1 General framework and quality dimensions

"Quality indicators are the most widely used tools to measure the quality of statistics. Indicators are developed to change the measurability level from nominal to ordinal or interval scale, and to find indicators more closely related to the phenomenon, namely to move from indirect to direct indicators. Quality indicators make the description of a statistical product more informative and increase transparency. The statistician or the user can assess the quality of different surveys or the same data in different periods by using the quality indicators. Indicators always simplify reality"[5].

However, sometimes users can misread quality indicators so Eurostat recommends to include qualitative statements helping them interpreting quality information in the wright way.

Quality is defined in the ISO 8402 - 1994[6] as: "the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs".

Eurostat defines quality of statistics according to the following six quality dimensions/criteria[7]:

- Relevance: the degree to which statistics meet current and potential users' needs. It refers to whether all statistics that are needed are produced and the extent to which concepts used (definitions, classifications etc.) reflects user needs.
- Accuracy: the closeness of computations or estimates to the exact or true values (a distinction is done between sampling and non-sampling errors).

[5] Eurostat. 2007. *Handbook on Data Quality Assessment Methods and Tools*. Wiesbaden.

[6] International Organization for Standardization: https://www.iso.org/standard/20115.html.

[7] Eurostat 2003 *Standard Quality Indicators, Producer-Oriented*. Working Group "Assessment of quality in statistics", Luxembourg, 2-3 October 2003.
Eurostat 2005. *Standard Quality Indicators*. Working Group "Quality in statistics", Luxembourg, 23-24 May 2005.

- Timeliness and punctuality: *Punctuality* refers to the time lag between the release date of data and the target date when it should have been delivered. *Timeliness* of information reflects the length of time between its availability and the event or phenomenon it describes.
- Accessibility and clarity: *Accessibility* refers to the physical conditions in which users can obtain data (where to go, how to order, delivery time, clear pricing policy, convenient marketing conditions, availability of micro or macro data, various formats, etc.). *Clarity* refers to the data's information environment, whether data are accompanied with appropriate metadata, illustrations such as graphs and maps, whether information on their quality also available and the extent to which additional assistance is provided by the NSI.
- Comparability: aims at measuring the impact of differences in applied statistical concepts and measurement tools/procedures when statistics are compared between geographical areas, non-geographical domains, or over time.
- Coherence of statistics is their adequacy to be reliably combined in different ways and for various uses. When originating from a single source, statistics are normally coherent in the sense that elementary results derived from the concerned survey can be reliably combined in numerous ways to produce more complex results. When originating from different sources, and in particular from statistical surveys of different nature and/or frequencies, statistics may not be completely coherent in the sense that they may be based on different approaches, classifications and methodological standards.

For each of the above dimensions a set of standard producer-oriented indicators can be defined to measure all quality components identified in the quality definition of Eurostat. These indicators should be easy to understand and should best describe the respective quality dimensions. Furthermore a decision to take refers to the more appropriate computation periodicity of quality indicators, for example, some of them could be produced in line with the frequency of production or dissemination of the output they refer to, others could be calculated once for longer periods. Undoubtedly their computation frequency depends on the purpose of each quality indicator (for example: monitoring quality over time) or on the survey or on the output dissemination frequency.

2.2 Key pilot quality indicators for SPPIs

The objective of the pilot example described in this paper is trying to measure the overall quality over time for SPPIs.

First of all, to better understand the example it is necessary to explain that, usually, comparisons among indicators can be done if indicators are first normalized and then standardized[8]. Normalization transforms indicators from absolute values to relative ones; standardization eliminates differences due to different units of measures (for example: units, square meters, etc.) and allows the comparison of the indicators.

The analysis proposed in this document can be summarized as a procedure done by steps.

[8] Cecconi C., Polidoro F., Ricci R. 2004. *Indicators to define a territorial quality profile for then Italian consumer price survey*. European Conference on Quality and Methodology in Official Statistics. Mainz. May 24-26.

The first step identifies a set of appropriate quality indicators to represent better each of the quality dimensions described by Eurostat. The second step normalizes only the indicators expressed in absolute values to transform them into relative values. In the case considered the totality of the indicators is expressed in units and can already assume values included in a 0%-100% range, for this reason no indicator needs to be standardized. The last step opportunely synthetizes indicators to provide an overall measure of quality for SPPIs.

The current proposal takes into account that the particular nature of economic indicators does not allow implementing all the key measures Eurostat proposes and, in particular, that SPPIs imply the use of non-standard methodology such as, for example, judgmental or cut-off sampling of units and typical estimation of missing prices. For these reasons, for five out of six quality dimensions, this work proposes in Table 1 a list of quality indicators for SPPIs chosen among the totality of measures that are theoretically able to represent each dimension.

Table 1 – List of pilot indicators chosen for SPPIs

Quality dimension	Indicator
Relevance	R1 - STS Regulation data completeness rate R2 - Series length completeness rate
Accuracy	A1 - Un-weighted unit response rate A2 - Weighted unit response rate A3 - Un-weighted item response rate
Timeliness and punctuality	TP1 - Time lag final results TP2 - Punctuality – delivery and publication
Accessibility and clarity	AC1 - Number of publications disseminated
Comparability	C1 - Length of comparable time-series C2 - Number of comparable time-series

The pilot example does not consider Coherence, the last dimension defined by Eurostat, because for SPPIs it is always fully satisfied. In fact the indices originate from a single source, represented from surveys collecting prices; therefore elementary results can be combined in numerous ways to produce more complex, but always, coherent results. For this reason no key measures have been described for this criterion.

As far as Accuracy is concerned, standard approaches aim to characterize sampling errors and non-sampling errors. However this scheme is hardly applicable to SPPIs especially because they are not obtained according to standard procedures and the standard errors of the indices are not known. Besides, for almost all cases, samplings of units are not probabilistic. For these reasons the pilot example considers Accuracy only from a non-sampling errors point of view.

Indicators in Table 1 are defined as follows:

1) *Indicator R1: STS Regulation data completeness rate.*
Target value: 100%, meaning that all the required series are produced.

$$\frac{number\ of\ SPPIs\ series\ produced}{number\ of\ SPPIs\ series\ required\ by\ STS\ Regulation} * 100$$

2) *Indicator R2 - Series length completeness rate*
Target value: 100%, meaning that all the required quarterly indices are produced.

$$\frac{number\ of\ quarters\ produced}{number\ of\ quarters\ required\ by\ STS\ Regulation} * 100$$

3) *Indicator A1 - Un-weighted unit response rate*
Target value: 100%, meaning that all the units (enterprises) in the sample give prices.

$$\frac{number\ of\ responding\ units\ in\ the\ sample}{total\ number\ of\ units\ in\ the\ sample} * 100$$

4) *Indicator A2 - Weighted unit response rate*
Target value: 100%, meaning that all the units (enterprises) in the sample, weighted with their own turnover, give prices.

$$\frac{sum\ of\ turnover\ of\ responding\ units\ in\ the\ sample}{total\ turnover\ of\ units\ in\ the\ sample} * 100$$

5) *Indicator A3 - Un-weighted item response rate*
It is the complementary index of the *Imputation Rate* (number of imputed prices in the sample/total number of prices in the sample) whose target value is 0% (no price is imputed).
Target value: 100%, meaning that all the prices in the sample are given by respondents.

$$\frac{number\ of\ prices\ collected\ in\ the\ sample}{total\ number\ of\ prices\ in\ the\ sample} * 100$$

6) *Indicator TP1 - Time lag final results*
Target value: 0% would be the ideal one; small values denote higher timeliness.

$$\frac{(release\ date\ of\ final\ results - \ last\ day\ of\ the\ reference\ period\ of\ SPPIs)}{number\ of\ days\ in\ the\ quarter\ of\ dissemination} * 100$$

7) *Indicator TP2 - Punctuality – delivery and publication*
Target value: 0% meaning that there is no delay on the delivery of data.

$$\frac{|actual\ date\ of\ dissemination - scheduled\ date\ of\ dissemination|}{number\ of\ days\ in\ the\ quarter\ of\ dissemination} * 100$$

8) Indicator AC1 - Number of publications disseminated

Target value: 100% meaning that SPPIs are published on every existing publication concerning producer prices.

$$\frac{number\ of\ publications\ on\ SPPIs}{number\ of\ publications\ on\ PPI} * 100$$

9) Indicator C1 - Length of comparable time-series

Target value: 100% meaning that the total of the quarterly SPPIs are comparable. Undoubtedly a long time series is desirable to conduct analysis without taking account of changes (in concepts, methodology, etc.).

$$\frac{number\ of\ comparable\ quarters\ for\ each\ serie}{total\ number\ of\ comparable\ expected\ quarters\ for\ all\ the\ series} * 100$$

10) Indicator C2 - Number of comparable time-series

Target value: 100% meaning that all the series are comparable. In this context "comparable" means that every quarter of each year is produced.

$$\frac{number\ of\ comparable\ series\ produced}{total\ number\ of\ comparable\ expected\ series} * 100$$

2.3 Pilot example and results

The pilot example refers to SPPIs, coming from direct surveys, computed from 2012 to 2016. For each quarterly service producer price index and for each year, annual averages of quality indicators values are calculated. Results are shown in Table 2.

Table 2 – Values of key pilot quality indicators for SPPIs from 2012 to 2016

QUALITY COMPONENT /NUMBER	INDICATOR	2012	2013	2014	2015	2016
R1	DATA COMPLETENESS RATE	70.6	70.6	94.1	100.0	100.0
R2	SERIES LENGHT COMPLETENESS RATE	70.6	70.6	91.2	100.0	100.0
A1	UN-WEIGHTED UNITS RESPONSE RATE	92.0	91.6	85.0	85.6	89.0
A2	WEIGHTED UNITS RESPONSE RATE	91.8	91.8	90.2	92.1	93.7
A3	UN-WEIGHTED ITEM RESPONSE RATE	95.5	93.4	93.3	90.8	89.9
TP1	TIME LAG FINAL RESULTS	94.6	94.0	92.9	93.4	92.3
TP2	PUNCTUALITY - DELIVERY AND PUBLICATION	0	0	0	0	0
AC1	NUMBER OF PUBLICATIONS	60.0	60.0	80.0	80.0	60.0
C1	LENGHT OF COMPARABLE TIME-SERIES	70.6	70.6	86.8	100.0	100.0
C2	NUMBER OF COMPARABLE TIME-SERIES	70.6	70.6	88.2	100.0	100.0

In order to provide a consistent and homogeneous interpretation of results, it is important to evaluate how to read indicators. In fact, taking into account that the target value of each indicator

is 100%, except for the two indicators that measure timeliness and punctuality (TP1, TP2), the percentage rates dynamic can be interpreted as a change in the quality of the price indices: if the value of the key indicator increases this means an increase in the quality of the corresponding dimension and vice versa. Indicator TP2 always gives the best results; in fact 0% means no delay on the delivery of data for each year, so in this case an increase of the value should mean a decrease in quality for the timeliness and punctuality dimension. The same interpretation should be given to TP1 indicator value (trends are inversely proportional). For this reason, Table 3 replaces both the Timeliness and Punctuality dimension key indicators with their complementary values (missing values up to 100%) that allow a homogenous comparison of the totality of the set of measures.

Table 3 – Homogeneous values of key pilot quality indicators for SPPIs from 2012 to 2016

QUALITY COMPONENT /NUMBER	INDICATOR	2012	2013	2014	2015	2016
R1	DATA COMPLETENESS RATE	70.6	70.6	94.1	100.0	100.0
R2	SERIES LENGHT COMPLETENESS RATE	70.6	70.6	91.2	100.0	100.0
A1	UN-WEIGHTED UNITS RESPONSE RATE	92.0	91.6	85.0	85.6	89.0
A2	WEIGHTED UNITS RESPONSE RATE	91.8	91.8	90.2	92.1	93.7
A3	UN-WEIGHTED ITEM RESPONSE RATE	95.5	93.4	93.3	90.8	89.9
TP1	"COMPLEMENTARY" TIME LAG FINAL RESULTS	5.4	6.0	7.1	6.6	7.7
TP2	"COMPLEMENTARY" PUNCTUALITY - DELIVERY AND PUBLICATION	100.0	100.0	100.0	100.0	100.0
AC1	NUMBER OF PUBLICATIONS	60.0	60.0	80.0	80.0	60.0
C1	LENGHT OF COMPARABLE TIME-SERIES	70.6	70.6	86.8	100.0	100.0
C2	NUMBER OF COMPARABLE TIME-SERIES	70.6	70.6	88.2	100.0	100.0

The increase in time of the value of most indicators can be interpreted as an improvement in the quality of the statistics produced. As regards non-sampling Accuracy, the response rate is good and the weighted key indicator (A2) increases despite the slight decrease of the un-weighted one (A1). Instead, better results can be surely obtained for the un-weighted item response rate (A3), for example, simply sensitizing respondents. Furthermore, some efforts should be done to improve the time lag of final results (TP1) shortening the time between the end of the reference period and the date of results. Finally, can be observed that time variability of the number of publications indicator (AC1) is high due to few numbers on the basis of which it is calculated.

The following averages of key indicator values are adopted to synthesize quality:

- *Simple arithmetic mean*

  It is the simplest way and is implicitly based on the assumption that different indicators are interchangeable and each of them is considered equally important (same weight). It eliminates subjective elements in estimating weights and allows a clear interpretation of results. However, it must be taken into account that interchangeability of measures is more likely if considered inside each dimension, among dimensions it could be less suitable. Furthermore it is known that simple arithmetic mean is affected by outlier, so it could not provide robust statistical results.

- *Geometric mean*
  Compared to the arithmetic mean, one of its characteristics is that small values are much more influential than the big ones. It is more sensitive to the variability of the components and therefore compensates for fewer indicators of different intensity.
- *Weighted arithmetic mean*
  It has the advantage of considering the indicators differently by assigning them weights but, at the same time, introducing subjective elements precisely in determining the weights themselves. Starting from the assumption that the best indicator is the one that assumes time values with the smaller distance from the ideal situation (target value), the weighting system is obtained on the basis of the distance of each indicator value from its target value. The greater the distance from the ideal situation, the more it is believed that the indicator should weigh in the synthesis.

In Table 4, the results coming from the three synthesis methods seems to stand for a general improvement of the average quality over time, from 2012 to 2015. The last year presents a small decrease due to the interruption of the press release in 2016, in fact indicator AC1 decreases (Table 3). However during 2017 a new press release is spread.

The methods give similar results thus indicating robustness in the measurements even though, as was expected, the use of weights has produced some subjective effect.

It would be interesting to repeat the exercise adding 2017 data and analyzing quarterly indicators instead of annual average ones.

Table 4 – Methods of synthesis and results

TYPE OF SYNTHESIS	2012	2013	2014	2015	2016
SIMPLE ARITHMETIC MEAN	72.7	72.5	81.6	85.5	84.0
GEOMETRIC MEAN	60.4	60.9	69.6	72.1	71.4
WEIGHTED ARITHMETIC MEAN	79.8	79.8	90.7	91.3	85.3
% VARIATION VALUE	**2012**	**2013**	**2014**	**2015**	**2016**
SIMPLE ARITHMETIC MEAN	-	-0.3	12.6	4.8	-1.8
GEOMETRIC MEAN	-	0.8	14.3	3.6	-1.0
WEIGHTED ARITHMETIC MEAN	-	0.0	13.7	0.7	-6.6

3 Conclusions

In recent year it is becoming more and more important measuring quality dimensions of statistics and analyzing interactions between the different quality components. Eurostat effort in this field has the purpose of making European statistics available and reliable as well as harmonized. For this reason European countries are asked to systematically produce reports giving metadata and quality indicators of their statistical outputs.

A lot of work must be done especially in estimating quality indicators for each quality dimension in the field of short term statistics where the time pressure to compile data does not help and where it would also be desirable to have standardized indicators included in statistical production processes themselves. Furthermore overall quality is a target even more difficult to

achieve and, above all, measure. The pilot example in this document is only an unpretentious attempt in this direction.

Bibliography

Cecconi C., Polidoro F., Ricci R. 2004. *Indicators to define a territorial quality profile for then Italian consumer price survey*. European Conference on Quality and Methodology in Official Statistics. Mainz. May 24-26.

Eurostat 2003 *Standard Quality Indicators, Producer-Oriented*. Working Group "Assessment of quality in statistics", Luxembourg, 2-3 October 2003.

Eurostat 2005. *Standard Quality Indicators*. Working Group "Quality in statistics", Luxembourg, 23-24 May 2005.

Eurostat. 2007*. Handbook on Data Quality Assessment Methods and Tools*. Wiesbaden.

Eurostat 2014. *ESS guidelines for the implementation of the ESS quality and performance indicators (QPI)*. Luxembourg.

Mazzi G.L., Museux J.M., Savio G. 2005. *Quality measures for economic indicators*. Eurostat. Working papers and studies.