Measuring the quality of multisource statistics
Sorina Văju (Sorina-carmen.vaju@ec.europa.eu), Mihaela Agafiţei, Fabrice Gras, Wim Kloek,
Fernando Reis

Keywords: quality, multiple sources, integration, administrative sources, big data.

1. INTRODUCTION

Most EU Member States have been moving towards an increased use of administrative data sources for statistical purposes, as a substitution and/or as a complement to survey data. At the same time, the emergence of big data allows for a further increase of available sources for statistics. Statisticians are looking for new ways to combine sources and methods in order to accommodate new demands for statistics. As a result, statistical output is based on complex combinations of sources. Its quality depends on the quality of the primary sources and the ways they are combined. This paper analyses the appropriateness of the current set of quality measures for multiple source statistics, explains the need for improvement and outlines directions for further work.

2. QUALITY OF STATISTICS IN A MULTISOURCE ENVIRONMENT–GENERAL DISCUSSION

The ESS quality framework identifies five quality dimensions to describe output quality: (a) relevance (European Statistics meet the needs of users), (b) accuracy and reliability (statistical outputs accurately and reliably portray reality), (c) timeliness and punctuality (statistical outputs is released in a timely and punctual manner) (d) coherence (statistical outputs are consistent internally, over time and comparable between regions and countries) and comparability (it is possible to combine and make joint use of related data from different sources), (e) accessibility and clarity (statistical outputs are presented in a clear and understandable form, available and accessible on an impartial basis with supporting metadata and guidance) [1].

Some quality dimensions – relevance, accessibility and clarity – are not impacted by integrating multiple sources while others – timeliness and punctuality – may be impacted but the way we measure them is still appropriate. However, measuring other dimensions – accuracy and reliability, coherence and comparability – require incorporating the effect of sources and integration approach. More specifically, the first two groups describe the statistical product irrespective of the statistical process behind it (the choice of data sources to use, the statistical processing and the integration approach). The last group highly depends on the quality of sources and of the way they are combined, as it focuses on measuring the deviations from reality and on indicating the correct use of the statistical product. At each step of the production process [2], accuracy and comparability appear as the quality dimensions that are actually at stake.

Even if some accuracy measures (e.g. coverage rate, edit failure rate, imputation rate, average size of revisions, etc.) can apply to the several types of data sources it is very difficult to assess the sensitivity of the final statistical output to source specific errors and to the methods used to integrate them. Consequently, the accuracy and reliability dimension needs to be reconsidered in order to cover all methodological aspects and implications given by the combination of sources and methods. International comparability can be seriously affected when integrated statistics include administrative data coming from different national administrative systems and produced using different

1 Eurostat, European Commission.
methodological approaches/combinations. At European level, this translates into a huge number of possible sources of lack of comparability, given by combinations of: (i) national legal and institutional environments, (ii) acceptable trade-off between quality dimensions at national level; (iii) appropriate trade-off between costs and benefits in terms of output data quality at national level, (iv) methodological choices to integrate the several data sources.

3. METHODS FOR QUALITY ASSESSMENT

There are three facets for which quality can be checked: input, process and output. Input assessment refers to the quality of raw data and should allow statisticians to decide whether and how a given data source – including big data and administrative sources – can be used on a regular basis to produce statistics. Process quality refers to intermediate steps; it describes or quantifies the transformations that the raw data has undergone through the statistical process (e.g. imputation, editing). Output quality refers to the final statistical product and it should provide to the user easy to understand information on the quality of the final data.

3.1. Output quality assessment on the basis of input and process

The natural approach for identifying the possible impact on quality of combining several types of data sources in the statistical production process is to look at each step of the production process and assess the impact of such integration. The use of combined sources mainly impacts the way the accuracy measurement is made. The assessment of the other quality dimensions does not specifically depend on using combined sources, with the exception of the comparability dimension. Nevertheless, comparability assessment can be to a large extent reduced to structural error generated by the introduction of some possible statistical biases. This does not affect comparability over time, for which the break in time series and the outliers are the main threats. Possible outliers/breaks can be detected based upon existing methods; this will, as illustrated later, provide some first insights on how to assess the quality of data derived from multiple sources.

Table 1 gives an overview, for several statistical production activities, of the link between the risks of combining multiple data sources and the corresponding impacted quality dimensions and quality measurement. Accuracy assessment of the combination of sources should most likely focus on aggregating random mechanisms effects with the bias effects introduced by non-survey data. However, when using multiple sources, measuring final data accuracy via assessment of the data integration in the several statistical production activities appears not straightforward and even too complicated to be envisaged.

<table>
<thead>
<tr>
<th>Statistical production activities</th>
<th>Risk</th>
<th>Impacted quality dimension</th>
<th>Error measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linkage and determination of the target population</td>
<td>Missed link, wrong link: under/over coverage</td>
<td>Accuracy, comparability</td>
<td>Bias, confidence range of the target population</td>
</tr>
<tr>
<td>Concept/definition</td>
<td>Aggregation of different concept/definitions</td>
<td>Relevance, accuracy, comparability</td>
<td>Bias, Variance error, qualitative assessment</td>
</tr>
</tbody>
</table>
### 3.2. Direct output quality assessment

In this section we discuss possibilities to assess accuracy and comparability of statistical outputs without analysing the processes behind it. There are three options: direct assessment of the output quality on the basis of the output itself, assessment on the basis of a common reference source and methods involving mainly bootstrapping techniques.

There are several ways to assess the output quality on the basis of the output itself. Breaks in series are a direct indication of bias and show the impact of changes in sources and methods. The impact can be measured by keeping for a while a double production system or by extrapolation. Bias can be indicated by systematic corrections when doing revision of data when more information becomes available. In case the revisions show systematic corrections, this would be an indication of bias. Another example is applying outlier detection techniques to cross-sectional data. The main advantages of the direct assessment methods using solely the output are: (i) they require no knowledge about the sources and methods used in the statistical production process and (ii) they are fairly easy to implement. The major disadvantages are: (i) it is not always possible to distinguish between real differences, bias and variation and (ii) the method offers no clue on diagnosis and remedy.

As regards the assessment of output quality with a common reference data source, two cases are distinguished: the quality survey and any other reference source. The advantages of the quality surveys are: (i) the quality survey has a known variance and it is designed to have a low bias; (ii) it can have diagnostic value by identifying the weaknesses of the process steps; (iii) it is easy to summarise into an overall assessment. In practice costs and other practical considerations will probably prevent its full scale application. At a less ambitious scale it might be possible to assess specific elements where other information is lacking (e.g. under-coverage). Other reference sources might be other related statistics, administrative sources or big data sources with considerable conceptual harmonisation. The advantages are: (i) low additional costs and no additional burden; (ii) the separate production process. The main disadvantages are: (i) for an administrative reference source, or in the case of big data sources, we have no control over variance and bias and thus it will often require an assumption on the level of variation and on the stability/equal distribution of bias; (ii) usually it has no diagnostic value; (iii) the natural tendency to incorporate good sources into the production process, thus making them unavailable as independent reference source.

The ESSnet AdminData [3] proposed ways to adapt the bootstrap re-sampling methods in order to estimate the root mean square error (RMSE) that includes both sampling variance and bias due to non-sampling errors, incorporating thus the effect of interaction with administrative data. The reasoning behind is that bootstrap methods enable inserting randomness through the replication of samples [4]. Thus, replications of combined dataset are produced, either by simulating the distribution followed by the data or by
using existing samples for replication. The purpose is to simulate and/or replicate random behaviour of administrative data by undertaking statistical inference on administrative data. These methods can equally be applied to big data sources.

Table 2. Bootstrap methods use by type of combination of administrative data with other sources

<table>
<thead>
<tr>
<th>Possible use</th>
<th>Remarks</th>
<th>Main practical problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replacement for primary and/or complementary data</td>
<td>Overlapping survey data can significantly increase the feasibility and relevance of the method</td>
<td>Inference on the distribution and/or generating process of the administrative data. Detection of break and outliers in time series.</td>
</tr>
<tr>
<td>Partial use for sample design or input for statistical registers</td>
<td>Uncertainty can be inserted by estimating false positive and negative probability</td>
<td>How to simulate the addition of a previously non selected unit in the replication of the sample</td>
</tr>
<tr>
<td>Additional variables for estimation; auxiliary information to support processing of primary data (editing, imputation, calibration)</td>
<td>Modelling on how random is channelled through the production process requires a good description of the production process</td>
<td>Simulation of the error caused by the imputation/estimation methods.</td>
</tr>
</tbody>
</table>

4. CONCLUSIONS

Measuring output quality through input and process quality gets too complex in processes combining several sources, especially at the European level. Therefore, alternative solutions should be found. The paper lists three alternative approaches that do not depend on the design of statistical process: (a) direct output assessment; (b) a common reference source; (c) bootstrapping.

Information on output quality has internal use for monitoring and improving the statistical production process. The information also has an external role. The quality information should be summarised in such a way that data users can assess the accuracy and comparability. The alternative approaches contribute to the quality assessment for internal purposes, but a coherent external summary of information remains difficult. Assessing quality is not for free. Knowledge on quality is also required to allocate scarce resources between improving quality and measuring quality.

REFERENCES


