DATA REVISION POLICY OF HCSO

Contents

1 Introduction

1.1 Revisions and their background

The two most common needs[1] users have regarding statistical data are timeliness (access the latest data as soon as possible) and accuracy (describe the observed phenomenon as closely as possible). As these requirements usually contradict each other – users would like to access the latest data too early, i.e. before all the information required to produce accurate and reliable estimates are available at the statistical office –, the statistical authorities often have to reconcile timeliness with accuracy.

In order to provide timely data as early as possible, provisional estimates subject to some degree of uncertainty due to the incomplete source data are calculated and released for various statistical domains. As better quality and quantity data are available with time, provisional results are replaced by more accurate estimates in the course of revision.

Statistical data may also be revised when there are changes in methods, concepts, definitions or in classifications. The goal of these revisions is to assure data quality and to ensure that statistics are in line with international conventions, modified legal/ methodological frameworks and user needs. Since these revisions can cause breaks in time series and distort data analysis, such modifications are rarely implemented (once in every 5 to 10 years), and introduced all at once.

1.2 Revisions in official statistics

A publicly available document – which is called revision policy – that contains the general principles of data revision makes the revision procedures more transparent and comprehensible to users, thereby helps to maintain the trust in official statistics and to further enhance the usability of statistics. Thus, the revision policy plays a vital role in communication with users of official statistics.

Besides the general revision policy, domain specific[2] revision policies also exist in the official statistics. The general policy acts as a frame of reference for elaborating these specialized policies, which contain more specific provisions or go further into details to account for specific framework conditions and different user needs related to the specific domains of statistics.

1.3 Scope of application

The general principles described in section 3 are applicable to all sets of statistics at HCSO. The principles are flexible enough to handle the data specificities as well as constrains of different subject matter statistics derived from legal acts in force and different

---

[1] Several quality components are distinguished in official statistics, please check the Quality policy of HCSO for more details (http://www.ksh.hu/docs/bemutatkozas/eng/minpol_web_eng.pdf).

[2] According to the Domains of Official Statistics framework, statistical domains are the outputs of statistical activities covering the complete statistical business processes that can be characterized by same concepts and methodologies.

methodologies applied during data processing. This ensures that domain specific revision policies are fully compliant with the general one.

1.4 Reference frameworks

In the preparation of the HCSO's revision policy, the following conceptual frameworks have been considered:

- the official statistical principles of the UN (http://www.ksh.hu/docs/szolgaltatasok/pressroom/news_2014/fundamental_principles.pdf)
- the European Statistics Code of Practice, which states the following criteria about data revision:
  - o Advance notice is given on major revisions or changes in methodologies." (indicator 6.6)
  - o "Revisions follow standard, well-established and transparent procedures." (indicator 8.6)
  - o "Revisions are regularly analyzed in order to improve statistical processes." (indicator 12.3)

  (http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF/5fa1ebc6-90bb-43fa-888f-dde032471e15)
- the "ESS guidelines on revision policy for PEEIs", which have been approved by the European Statistical Committee in February, 2012
  (http://ec.europa.eu/eurostat/documents/3859598/5935517/KS-RA-13-016-EN.PDF)
- the quality guidelines of the HCSO
  (http://www.ksh.hu/docs/bemutatkozas/eng/minosegi_iranyelvek_eng.pdf)
- the dissemination policy of the HCSO
  (http://www.ksh.hu/docs/bemutatkozas/eng/tajpol_web_eng.pdf)
- finally, the HCSO takes into consideration the deadlines of mandatory international data transmissions and publications

2 Revisions

2.1 Definition of revision

A revision in official statistics is defined as any change in a value of a statistic released to the public. The fundamental aim of revision is to improve data quality and thereby to be more accurate when reflecting the reality. Therefore, revision is inevitable when the published data do not contain all the information necessary to precisely describe the phenomenon of interest. Thus, revision is a recognized and vital part of the statistical business process.

2.2 Reasons for revisions

Revisions take place for a number of distinct reasons. In reality, some of the distinctions are blurred because two or more kinds of revisions (derived from different reasons) may be made at the same time. The reasons tend to break into four groups:

Incorporation of better source data
a) incorporation of source data with more complete or otherwise better reporting
b) incorporation of source data that more closely match the concepts
c) replacement of first/provisional estimates based on expert judgments or on values derived by statistical techniques, or as a result of benchmarking (see page 4 for details)

Capturing routine recalculations
d) incorporation of updated seasonal factors
e) updating of the base period

Reflection of improved methodology
f) changes in statistical methods
g) changes in concepts, definitions and classifications

Corrections of errors
h) correction of errors in source data and computations

Occurrence of reasons in practice:

Incorporation of source data with more complete or otherwise better reporting
a) The first reason, **incorporating of source data with more complete reporting**, causes revisions across a wide spectrum of statistics. For example, the first/ provisional estimate of the monthly volume index of industrial production is released approximately 36–40 days after the end of the reference month (the estimate is provisional because it may be based on information available at that time, which usually incorporate only the data of large businesses with advanced reporting system). This means that the index is available very early, which enables a timely and up-to-date evaluation of industrial development. Few weeks later revised volume indices are calculated (for the same month) based on late or corrected incoming production reports, that is, these indices incorporate more and better quality source data, thus reflect the performance of industry more accurately.

Updating weights, such as those for calculating price indices, also belongs to this class of reasons. This brings in information from more recent surveys, thus improves the quality of estimates.

b) The second reason, **incorporation of source data that more closely match the concept**, is most likely to occur in datasets that put together many data sources (e.g. national accounts), representing a comprehensive picture of some aspect of the economy. For example, if production is to be measured, source data that represent sales may provide a first estimate. As data more closely matching production become available, the initial estimates are revised.

c) **Replacement of first/provisional estimates based on expert judgements or on values derived by statistical techniques** as a reason for revision occurs in situations when no current data may be available, and the first estimates are based on judgments or on statistical techniques (e.g. on models). A revision may result when data become available.

   Benchmarking means adjusting higher frequency (e.g. monthly or quarterly) data – which typically based on a smaller sample – to the corresponding lower frequency (e.g. yearly) results, based on more complete, more reliable and higher quality information which become available only later. Benchmarking can be applied when an indicator is available at different frequencies (e.g. in case of railway transport, HCSO collects data on quarterly and on yearly basis as well).

Capturing routine recalculations

d) The **incorporation of updated seasonal factors** relates closely to the first group (incorporation of additional source data), in some cases they are not identified as two separate reasons due to the nature of seasonal adjustment. Seasonal factors can change as the new time series value comes into the calculation, which may result in the alteration of the whole time series.[3]

e) The **updating of the base period** of an index – that is, the year set equal to 100% – is often a reason for revision. The base year is rarely refreshed, typically in every 5 or 10 years.

Reflection of improved methodology

f) The sixth reason, **changes in statistical methods** often go hand in hand with changes in source data. However, they can also occur independently. For example, revision analysis may reveal that improvement in accuracy or timeliness can be achieved by changing or improving the current methods.

g) **Changes in concepts, definitions and classifications** – which may also lead to revision – often stimulated by the adoption of new international guidelines, standards or new legal frameworks. For example, the change in Statistical Classification of Economic Activities (NACE[4]) which took place in 2008, was due to a new regulation.

Methodological changes occur less frequently than the previous reasons (only every 5 to 10 years). If no revision would be carried out when such changes happen (i.e. the time series of indicators and components would not be recalculated according to new standards, classifications, etc.), breaks in time series would occur, which would restrict the comparison of different time periods. This could easily lead to misunderstanding. An example for

---

[3] For more information on seasonal adjustment, please check the following link: http://www.ksh.hu/docs/eng/xftp/modsz/eszezonalis_kiig.pdf

[4] Detailed description about the change in NACE can be found in the following document: http://www.ksh.hu/docs/osztalyozasok/teaor/teaor_methodology.pdf

change in methodology is the adoption of the new EU accounting framework, the European System of National and Regional Accounts in 2010 (ESA 2010[5]).

Correction of errors

    h) Finally, revisions occur as **correction of errors in source data and computations** is carried out. Errors may appear in source data (for example, a reporting unit may discover after submitting the data that some components are missing), or in any step of the statistical business process.

2.3 Classification of revisions

Taking into account the various causes of revisions and the different frequencies of publications, the Hungarian Central Statistical Office – in correspondence with international guidelines – distinguishes the following types of revisions:

1 Planned revisions

Planned revisions occur regularly based on a predetermined schedule and are pre-announced in the revision calendar available on the HCSO's official website. Considering the degree of the necessary methodological adjustments (if any) and the length of the revised time series, the planned revisions are divided into two groups:

- **Routine revisions**

Routine revisions are changes in published data which are related to the regular statistical business process. Routine revisions mainly occur when the incorporation of late information (new or the correction of already obtained) modifies the already published results or in the case of benchmarking. For example, the replacement of estimated values of missing information with actual data received after the deadline or the substitution of provisional results with final results when better quality and quantity data are available also lead to routine revisions.

Routine revisions stem from the first four reasons: incorporation of source data with more complete or otherwise better reporting (a-c) and incorporation of updated seasonal factors (d).

Routine revisions are conducted periodically, according to a schedule drawn up in advance. With some sets of statistics (e.g. in foreign trade), a number of revisions are needed to obtain final results, while in other cases the provisional results are replaced by final data in the course of one single revision.

Routine revisions barely affect the applied methodology, and only a few periods[6] (some months or quarters) back in time are revised and longer revisions take place at a lower frequency, e.g. annually.

- **Major/Methodological revisions**

Major revisions are changes in published data, often substantial, which are due to changes in definitions, classifications and methodologies. New concepts, definitions

---

[5] For more information on the change from ESA 95 to ESA 2010, please visit the following website: http://ec.euro-pa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-14-002?inheritRedirect=true&redirect=%2Feuros-tat%2Fweb%2Fesa-2010%2Fmanuals-guidelines

[6] In case of seasonal adjustment, the whole time series is revised.

and classifications introduced to meet new international standards, updating of the weights of the base year of an index series, the availability of a new structural source that is only collected at long intervals (5 to 10 years), such as the census, and the entry into force of a new legal act may also cause major revisions. To sum up, item e and g from the above list are responsible for major revisions.

The introduction of the new concepts of the European System of National and Regional Accounts (ESA) 2010 in 2014 September, which replaced the ESA 1995, is considered as a major revision. The ESA 2010 modernized the general standards, classifications and accounting methods used to create national accounts.

Major revisions are planned very well in advance and users should be informed beforehand on the forthcoming major revisions. They are less frequent than routine revisions and occur only every 5 to 10 years. Since major revisions affect a large part of the time series and sometimes even the complete time series, it is necessary to back-cast time series, otherwise major revisions would produce breaks and inconsistencies in them. The length of the revised time period depends on user needs, costs and available resources.

As major revisions may have significant impact on already published data, the statistical office tries to concentrate and coordinate such revisions in order to avoid the negative effects of too regular revisions.

2 Unplanned revisions

Unplanned or unscheduled revisions are those that are not foreseen (as opposed to planned revisions), because they are a result of unforeseeable events and therefore it is usually not possible to pre-announced them in advance. Unscheduled revisions are not part of the regular statistical business process. They appear due to significant errors detected after statistical indicators were already published (item h from the above list), unpredictable changes in methodology, unpredictable emergence of new and better quality data, unpredictable changes in reported or administrative data and technical problems arising during data processing.

As unscheduled revisions can undermine confidence in the quality of official statistics, HCSO is committed to avoiding as much as possible unscheduled revisions and to limit them to the case of important errors (whose correction results in significant improvement regarding data quality). Unscheduled revisions are communicated to the users in a transparent manner

3 Principles for revisions

HCSO takes into consideration the following principles for revisions:

- HCSO makes its general and domain specific revision policies publicly available on its official website.
- HCSO applies these principles to all of its statistics. Although subject matter statistics may have specificities regarding their revision practices, they have to be fully compliant with the general principles.
- When defining domain specific revision policies, HCSO strives to reach as much consistency as possible not only internally (between yearly and quarterly/monthly estimates of the same subject matter statistics), but across different statistical domains.
  - o In case of routine revisions, the frequency and timing of revisions, and the length of the revised time series should be synchronized for related statistical domains. This will improve their comparability and facilitates the interpretation of statistical data for users.
  - o Dates of revision and how far the series are back-casted should be coordinated in case of major revisions to make the time series of related statistics comparable and consistent.
- To foster user confidence and the transparency of the revision processes, HCSO aims at maintaining the stability of general and domain specific revision policies over time. A stable revision policy ensures that users generally know in advance when, which and why data will be revised.
- HCSO notifies users about forthcoming revisions in time, and indicates their date and time in the revision calendar.
  - o Routine revisions occur regularly and at fixed dates, often together with the release of new provisional estimates (e.g. the volume indices of industrial production for the first quarter are revised when HCSO publishes provisional estimates for the second quarter). Users should be informed in advance about the main parameters of revisions, namely their frequency, date, depth (only aggregates are revised, or their components as well), length horizon (how many periods back in time are revised, e.g. only the previous quarter or the whole current year), and where it is possible (in case of key indicators), about the possible effects of revisions. The office documents these information and make them available for the public on its website (in the form of quality reports, methodological documentations, etc.).
  - o Major revisions are planned very well in advance and the date of their implementation are released to the public on the website of the HCSO or via other communication channels. The announcement briefly sums up the reason for the revision, the set of statistics involved, the release date of the revised data, the depth of revision, the length of the revised period and the foreseen impact on data. The publication of the revised data should be accompanied by a documentation which allow users to

understand and assess the revised results. This documentation, which is published either on the website of the HCSO or in a form of printed or online articles, quality reports, etc., details the reasons for revision, analyzes its impact on the aggregated data and if any breaks occur due to the revision in time series, it gives an explanation for that. If the major revision concerns several statistics and requires some time to implement, HCSO should notify users about the expected schedule.

- o As soon as it becomes clear that unscheduled revision is required, if it is possible, the users should be notified in advance via an announcement on the website of HCSO. With the publication of the revised data, a background note should be issued, which provides details on the reasons for the unscheduled revision, on the actions taken to correct the errors and an assessment of its impact on the published data.

- In order to make the revised data easily distinguishable from the non-revised ones, HCSO uses an adequate marking system in its publications. In addition, HCSO indicates the fact of revision in a textual note in its publication when it is possible. It should be clear from the marking system which data are provisional, which ones are revised and which data can be considered as final.

- When defining revision practices, HCSO tries to keep a balance between the additional information gained from revisions and the effort they cause to the users. The benefit that revised data offer should exceed the extra work it induces to users (who may have to adjust their own data bases, calculations, forecasts, etc.). Too frequent revisions, which provide only little additional information may reduce the user's trust in official statistics. Therefore, the burden caused by revisions should be continuously monitored and revisions should only be applied when the information gained exceeds a given threshold.

- o In case of routine revisions, the revision parameters and the publication dates should be determined in such a way that on the one hand, the first/provisional estimates provide sufficient quality information to the users, while on the other hand, the additional information gained through revisions surpasses the extra work occurring on the users' side.

- o As major revisions have more complex and bigger impact on data, implementing them too frequently can confuse data users. Hence, they should be carried out once in every 5 to 10 years and should be introduced all at once. This helps avoiding unnecessary breaks in time series, while at the same time it allows us to track the structural changes of the economy, social processes and to keep up with the best international practices. In addition, major revisions of different statistical domains should be synchronized and coordinated in order to maintain the comparability of their time series.

- o Due to their ad-hoc nature, unscheduled revisions may confuse users, thus damage their confidence in official statistics. Therefore, HCSO tries to avoid them and only implement unscheduled revision after careful consideration (it may be possible to integrate such revision into a forthcoming routine or major revision), taking into

account the improvement in data quality and the extra work it causes on the user side. If the correction of a significant error necessitates an unscheduled revision, it should be implemented as soon as possible in a transparent manner and should be announced to the public.

- In case of key indicators, HCSO monitors the impact of revisions on data. To this end, HCSO stores the real-time data tables (which captures the different historical versions of published data) in its archives, and uses them to conduct revision analyses. These help the office to optimize the revision processes, their parameters, the publication dates and to improve data quality. During the analysis, with the comparison of provisional and final estimates, we can measure the size of revisions (the degree of modification), assess the reliability of provisional estimates and check for systematic bias in first estimates. Furthermore, it can indicate flaws in the data collection and processing that may necessitate changes in the procedures. By examining these flaws in detail, searching for reasons and remedies, the data quality can be monitored and improved. In addition, based on revision analyses performed on previous periods, the likely size of future revisions can be assessed, which is quite important when provisional results are used by policymakers, investors or researchers.

HCSO publishes the revision practices of the domain specific statistics in their methodological documentation[7] following a clear, standardized structure that facilitates their understanding.

---

[7] http://www.ksh.hu/apps/meta.main?p_lang=EN