32nd Meeting of the Voorburg Group on Service Statistics

Notes on discussion at the Quality Indicators Poster Session

Participants:

Marcus Fridén, Rohan Draper (Sweden)

Cristina Cecconi, Fabrizio Marinucci (Italy)

Mary Beth Garneau (Canada)

Dr. Jutta Oertel, Dorothee Blang (Germany)

Bonnie Murphy (US BLS)

Discussion:

Index Auditing Model, Sweden

Statistics Sweden presented a demo of the Index Auditing Model as a poster session. Primary purpose of the Model is to understand where index reviews are likely needed (evidence-based), to provide a transparent consistent basis for resource allocation/planning and to give input to internal users own auditing (for example, national accounts). Secondary purpose was to increase the understanding of price indices (and their quality) among users and to increase own understanding of the types of process data that can be collected and analysed.

Questions received from VG members with answers:

- Q. Do you know what method of QA was used? A. Yes, the Swedish compilation system has been updated to allow analysts to select which QA method they have applied on a per specification/per reporting period basis.
- Q. Are ratings adjusted based on the industry being assessed? A. Yes, if an index has initially received a high risk rating, an investigation is conducted. For example, list prices being highly utilised would give a high risk result. If the industry was postal services, where list prices represent actual transaction prices, the rating would be adjusted accordingly.
- Q. How often are audits conducted? A. Audits are conducted annually at the two-digit level. If issues are discovered, audits can be conducted at the finest digit level.
- Q. Are categories weighted together to get a total score at the index level? A. Not at this time. Our intention is not to weight categories together at the index level as we do not want to mask any issues in a particular area. A risk is a risk.
- Q. Have the audits triggered any actions? A. Yes, for example, for one particular industry a specification review was implemented and staff received specification training.

Feedback

The project sounds very promising from the NA point of view. The results can be used in many ways:
- balancing supply and use tables
- when doing adjustments to current price figures in light of volumes
- when choosing the best index as a deflator
- preferably to introduce it in the IT system of NA's SUT and deflation IT system in Finland as well!

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

Statistics Italy presented a proposal to outline quality for SPPIs focusing on the quality of the statistical output choosing an ad hoc set of key indicators.
In Italy, the development of a set of quality indicators for short-term statistics (STS) is regularly requested by Eurostat, who systematically monitors and measures the quality of principal European economic indicators (PEEIs), which also include services producer price indices (SPPIs).
A global data quality assessment should be done according to three quality aspects, closely interrelated:
- characteristics of the statistical production process
- characteristics of the output (statistical product)
- perception of the statistical product quality from the users' perspective.

Eurostat focuses on the characteristics of the output (statistical product) and defines six quality dimensions: relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability, coherence. A set of quality indicators can be identified for each quality dimension.

A proposal is to try to measure the overall quality over time of SPPIs surveys.

The example (on SPPIs, coming from direct surveys and different industries, computed from 2012 to 2016)

- Identification of a set of QIs
- Normalization: to transform QIs into relative values
- Standardization: to cancel differences in units of measure (all QIs included in
  0   – 100% range)
- Synthesis of QIs: to compute averages of QIs (by calculating simple arithmetic mean/ geometric mean/weighted arithmetic mean)
- How to choose weights for Qis in the weighted arithmetic mean? Decision: the best indicator is the one that assumes, for each year, values with the smallest distance from the ideal situation (target value), namely the greater the distance from the ideal situation, the more the indicator should weight in the synthesis (weights = distance).

Comparisons among QIs coming from different industries or from different price statistics can be done because QIs are normalized and standardized, so they can be used in referrence with the same scale and unit of measure and can also be synthesized if needed.

Quality Assessment for Price Indexes: Performance Measure Grading Scheme, Canada

Statistics Canada presented a Performance Measure Grading Scheme to evaluate producer price indices index on key performance indicators to promote sound methodological practices and to convey overall quality and reliability of published index numbers. Its components were drawn from the OECD Generic Statistical Business Process Model and Statistics Canada's six dimensions of quality. Assessing the quality of an index is multi-faceted because of complexities of index numbers and calculations as well as different components of index compilation. An index number is comprised of price relatives, weights and a variety of forms of treatments of these data. The quality of an index must be assessed on individual parts as well as the whole. The grading scheme intends to capture this and provide a measure of quality for the entire index as well as its individual components, starting from a qualitative conceptual assessment to a quantitative processing perspective. The diversity of Canada's producer price index methods brings with it a complexity when developing a standard approach to assess data quality.

VG participants were asked for input and feedback on:

- any quality indicators that could be added;
- challenges in measuring variance;
- better measures of bias;
- how to weight and score various indicators;
- approaches to identify fitness for use of subindices in lower levels of the index aggregation structure

There are trade-offs to be made in programmes that would affect the scores of individual quality indicators. For example, a high score for timeliness could come at the expense of larger revisions. It was noted that weights would differ by use as one user might want data as early as possible, while another may favour accuracy and fewer revisions. Instead of a grading scheme, the suite of indicators might work better as a dashboard to identify the strengths and weaknesses of an index, leaving management to make decisions based on key use of an index (deflation, monetary policy, forecasting, contract escalation, etc.).

Some NSOs who currently calculate variance are not publishing the measures. Statistics Canada's Price Measurement Advisory Committee noted that bias is more important than CVs, since you could pick a biased sample that gives you very low variance.

One participant cautioned about publishing a quality grade for fitness of use of subindices. Her NSO publishes a letter grade based on only one dimension of the quality. This can sometimes cause confusion among users, since quality has many aspects that are hard to show in a single number.

Overall, many areas are investigating similar measures providing a great opportunity for collaboration. The accompanying papers prepared by various presenters in the session will be a useful resource for others who are looking to implement similar measures.

Quality control in a multiple-source mixed mode statistics, Germany

In 2007, Germany switched from a traditional survey to a multiple-source mixed mode system ("Mixmodell") in producing quarterly turnover indices for the service sector. This method is output-oriented. It successfully combines survey data for large enterprises with administrative value added tax data – relieving small and medium-sized enterprises of bureaucratic burden and saving costs for statistical offices. After methodological, organisational and processing problems have been solved, the focus lies on the challenge of controlling the quality in the "Mixmodell". Well-established quality indicators only partly meet the relevant challenges.

The relevant survey covers 40 – 60% of turnover and takes precedence over the use of administrative data. Only units that provide data *from the same source* for both the reference quarter and the previous quarter are included in the index calculation ("pairing principle"). Destatis improves the data quality to fulfil short-term statistical needs by applying editing procedures and using estimates and additional information from the business register. Owing to a legislation amendment concerning the use of administrative data for statistical purposes, Destatis and the statistical offices of the *Länder* are now authorised to clarify inconsistencies in the turnover tax data directly with the enterprises and to correct the data accordingly.

The implementation of the multiple-source mixed mode system resulted in a clear reduction of the statistical reporting duties of small and medium-sized enterprises. At the same time, the combination of survey and tax data produces a nearly complete count of the enterprise population. This makes a good case for the data quality.

Conclusions and feedback

Quality control in the "Mixmodell" focuses on the following principles:

- For primary survey data: increasing input and output quality
- For administrative data: monitoring input, automatic improvement of input and controlling output quality
- For data processing and final results: ensuring process control, completeness and timely delivery of data.

Quality control of primary data and processes can be described and ensured by means of well-established, well-tried methods and indicators. However, all but impossible input control with regard to administrative data is a permanent challenge. Relevance and accuracy of the administrative and finally also mixed data are hard to assess and cannot be ensured. Timeliness and accessibility of results can be largely ensured and documented, while controlling comparability and coherence has its limitations.

In order to improve quality control in multiple-source mixed mode statistics, the following measures are suggested:

- Statistical requirements should be taken into account in administrative data collection and processing. This is even more important when looking at future options of alternative (digital) data sources.
- The continual improvement of the method and the process is a "must".

- Quality indicators for administrative data and especially for mixed data should be further developed and standardised.
- Users should be enabled to assess the quality of multiple source mixed mode statistics and to understand their qualities and deficiencies, even if there are no standard instruments.

All in all, as a result of a wider access to alternative data sources, the statistical offices' responsibility to ensure and document the quality of the resulting manifold multiple-source mixed mode statistics will be a great challenge.

Variance estimates for price changes in the Producer Price Index, US BLS

The United States presented a poster describing variance calculations for SPPIs. Only a handful of countries reported performing variance calculations. The US is using a bootstrap method that treats the collected sample as the population and draws replicates, with replacement, from the original sample. Within each industry, variance strata are defined such that replicates are formed for certainty units and probability units separately. This method was chosen because the SPPIs use a two-stage sampling process with probability proportionate to size sampling down to the item level. One-month and 12-month median absolute percent change and median standard errors for NAICS indices down to the 6-digit industry level are published annually. The variances that were recently published for the first time are still under analysis. The variances are indicators of potential issues and not absolute quality measures. The same processing system is used for production and variance calculation, so the variances have lagged availability.

One of the main points in the discussion was the cost of developing the variance calculation within the production system. Although this ensures a very accurate variance calculation, it was suggested that a time and cost savings might be realized if variance calculations were done using SAS. We also discussed whether the concentration within the industry affects the variance. Concentration can affect variance calculations, so it is best not to compare across dissimilar industries (particularly concentrated versus non-concentrated). Variance is used more for following trends in variance for each industry. We also discussed what actions should be taken when an area has a large variance. Sample sizes could be increased, resampling could be expedited, and a search for alternative data to supplement the sample could be researched. In addition, the US does have a formula that incorporates variance into determining the sample size of industries.

The US has just recently begun to develop an "Index Quality Compass"[1] to measure the current health of each industry index. The variables that are measured in this tool are attached.

Delegates wanted to know what weights are assigned to each of the variables. The US SPPI does assign weights to many of the indicators, but the weights are dependent on what the analysis is being used for. This is a tool that was very recently developed and it is for internal use only (i.e. it is not published). There was also a suggestion to add a quality indicator variable that compares the output trend for the industry to the price trend.

---

[1] See Appendix – Index Quality Compass, US

Appendix – Index Quality Compass, US

Index Quality Compass

BLS
U.S. BUREAU OF LABOR STATISTICS

www.bls.gov

Variables Used For Index Quality Measurement:

- NAICS Industry
- Industry Title
- Index Age
- Survey Observations
- Establishments Surveyed
- Initiation Weighted Response
- Survey Weighted Response
- Historical Sector Benchmark of Weighted Response by Index Age
- Industry Representation Percentage

- Remaining Establishment Percentage
- Observation Percentage with Base Price Change
- Months Publication Failed
- Third Party Data Comparison Available
- Count of Publication Index Cells
- Average Observations per Index Cell
- Average Web Hits
- Median Absolute Variance
- Final-Demand Relative Weight

Index Quality Score:
Scores are based on variable distribution thresholds and weighted to reflect each variable's importance to index quality. Indices are penalized for variables outside warning thresholds and are used to calculate each index's score. Variables are also penalized more based on a greater distance outside a warning threshold. Index quality is measured on a scale of 0 to 100.

Index Quality Ratio:
Ratios are based on the weighted average of an index's variables to warning thresholds. Indices are rewarded for variables inside warning thresholds and penalized for variables outside. The greater the distance inside or outside warning thresholds, the greater the reward or penalty will be to an index. This holistic approach to index quality is measured from 0 to 1.