ENBES Workshop, online, 14-17 September 2021: Adapting establishment statistics to new conditions

Small group discussions

Topic 1: Data collection in relation to timeliness of output (in COVID times).

The group comprised members from several National Statistical Institutes and other institutes involved in Business statistics and focused on challenges and experiences in the area of data collection during the first 1 ½ year of the COVID pandemic.

1. A change in mentality at NSIs occurred in several ways:
   - act more quickly at NSIs, e.g. by adapting surveys to the changed circumstances, setting up new surveys, develop new types of output and increase timeliness and sometimes even shift to e.g. weekly instead of monthly data. And also by making use of restrictions for the better, e.g. testing new surveys with video meetings allows to recruit test persons from regions that would normally be too far away.
   - assume a different role vis à vis respondents and society as a whole, by proving that NSIs can quickly adapt in order to provide important and reliable information to enable society to handle a huge crisis.
2. Respondents are an asset to NSIs: 'Respondent Capital Management' is important.
   - Apply a more friendly approach and 'tone of voice' in communication with respondents, aimed at structural collaboration.
   - Don't emphasize that certain surveys are mandatory, but put the emphasis on the importance of the outcomes for government and businesses themselves.
   - Develop publications and applications that have value for the responding firms (like the web publication from Serbia).
   - Send first aggregated results to non-respondents as a means to remind them of filling out the survey and to motivate them.
3. Existing data are also an asset for NSIs, to further explore and bring to value. E.g., by combining data sets, providing more regional detail and increasing timeliness for more aggregated data.
4. Make more and better use of administrative data sources, for economic as well as for social statistics. Often done in combination with survey data.
5. Use new sources for existing statistics, like administrative data, surveys by email and online web questionnaires of paper surveys and web scraping. Also new private data sources may be used, such as aggregated data on payments (like bank or Payment Service Provider data) or mobility (using mobile phone data). This often works best when combined with existing data.
6. Strategic thinking is important and NSIs should prepare to cope with the unexpected: organize resilience and apply risk management.
   - Invest in more general and flexible statistical processes and IT-systems.
   - Maintain enough 'research capacity' to be able to cope with unexpected circumstances.
   - Carry out risk assessment and develop counter measures to monitor risks and to be able to act when a risk actually occurs.

Topic 2: Handling selectivity in available data sets/use of non-probability samples.

The group consisted of about five-six people from a few National Statistical Institutes. Our group examined the use of non-probability samples and how to handle selectivity within these. Our discussion centered around four main points.

(i) Discussing the possible use of online/big data sources for business statistics. Here we discussed how areas such as the Consumer Prices use web scraped and scanner data, either to augment or replace current/previous method. We discussed the use of admin/big data such as VAT turnover, COVID locations and population surveys to support ongoing business statistics. We briefly discussed opt-in surveys, but these were not used in a business capacity by the members of the group.

(ii) How to link non-random samples with registers. We discussed issues around linking, where there is no unique identifier available, e.g. name linkages, trying to link urls via web-scraping, etc. and where there are problems with subpopulations of the data.

(iii) Estimation techniques for non-random sample. Here we considered potential options and constraints in how to use background variables to correct for selectivity. We discussed whether a lack of (assessment of) representativeness can be outweighed by reduced costs, absence of response burden, lower measurement error and improved timeliness.

(iv) How to align information from multiple sources. We focused on the issue of measurement errors in terms of the misclassification of units, as well as target variables. For both types of errors, the need to understand the sources, combined with being able to distinguish any systematic versus random errors is crucial.

While we felt the use of non-probability samples is beneficial, our key takeaway was that this sort of data should be used as an auxiliary source/variable, rather than for producing official statistics or determining populations. It should be clearly stated in any publication, if non-probability samples are used, so the user is aware of any potential measurement errors and selectivity.

Topic 3: Regionalisation of statistics

The group consisted of five people from a few National Statistical Institutes. The context of the discussion was increasing user demand for statistics at more disaggregated levels, geographically and by classification. Approaches included:

- Small area estimation – Canada used a Fay-Herriot model to increase the number of Central Metropolitan Areas (sub-provincial areas) for which monthly statistics were good enough quality to publish. This approach was possible due to the availability of high quality admin data (GST) - which comprised monthly returns for large businesses, quarterly for medium and annual for small (the annual returns were calenderised to monthly using the seasonal patterns from the larger sub-annual returns).
- Modelling admin sources – the US had been using a state-space modelling approach for some time to estimate unemployment, similarly the Netherlands and the UK.
- Combining admin sources – Ireland were meeting the demand for county-level data by combining available sources. Some of which were COVID related and hence would not be available forever, so using VAT was being considered but there were issues to be overcome (estimates of turnover, balancing adjustments etc.).
- Direct use of VAT – the UK had made good use of VAT data, overcoming many of the issues faced by Ireland, in regional GDP. Direct use in surveys was more of a challenge, and this was still an area of research.

The consensus of the participants that more indirect use of admin data should be considered – whether modelling, or in use as auxiliary variables.

Other challenges discussed were:

- Ensuring consistency between low and high level estimates through direct/indirect estimation – a software solution using optimisation techniques is possible here, for example the UK uses FCOP-Xpress.
- Geography of statistics based on people was changing due to the increase of working from home.
- Geography of statistics based on businesses with multiple sites continued to be a challenge in terms of apportioning 'head office' values to regions, having up-to-date records, and there was a definitional issue here too – users needed to be clear on what they wanted and what they were getting!