32nd Meeting of the Voorburg Group on Service Statistics

New Delhi - India

October 23rd to October 27th 2017

Abstract for Poster Presentation:

Quality control in a multiple-source mixed mode statistics – using the example of German turnover indicators in the service sector

D▮▮STATIS
wissen. nutzen.

Dr. Jutta Oertel (jutta.oertel@destatis.de)
Statistisches Bundesamt – Federal Statistical Office of Germany

Table of contents

1. Introduction

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 the statistical offices.

Now that 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. This paper illustrates the experience gained with multiple-source mixed mode quality control in Germany.

2. Method of the "Mixmodell"

German short-term statistics on turnover in the service sector comprise sections H and J as well as large parts of sections M and N of the European Statistical Classification of Economic Activities (NACE Rev. 2). They are produced in accordance with the European regulations on short-term statistics.[1] In short-term statistics turnover is measured in the form of indices, currently based on 2010. The weights for producing the indicators in Germany are derived from the annual structural business survey in the service sector. The indices provide trend rather than level information. Preliminary results are available 60 days after the end of the reference quarter, revised results 90 days later and final results 180 days after the first publication. The indices are made available to the public in an unadjusted form, in a working-day adjusted form and in a seasonally adjusted form. From the first quarter of 2003 to the first quarter of 2007, the quarterly indices of turnover in short-term statistics were produced with the help of a 7.5% stratified random sample survey. Since the second quarter of 2007, turnover indicators in short-term statistics have been obtained with the help of a multiple-source mixed mode method. However, as revision 2 of NACE resulted in extensive classification changes for the whole service sector, data series for periods before 2008 are of limited informative value and comparability.

The multiple-source mixed mode method combines a primary survey conducted among large enterprises (i. e. with an annual turnover of at least 15 million euros and/or with at least 250 persons employed) with administrative data for small and medium-sized enterprises. Enterprises are obliged by national law[2] to provide information (online). Administrative data on turnover are provided by the tax authorities of the 16 German states *(Länder)*, the data stemming from turnover tax prepayment notices. Monthly turnover tax prepayment notices are obligatory for the full population with the exemption of very small units. A law on administrative data use for statistical purposes explicitly specifies its terms and conditions.[3] The data delivery process of the administrative data is established as a reliable routine.

---

[1] *Regulation (EC) No 1165/98 and Regulation (EC) No 1158/2005 of the European Parliament and of the Council*

[2] *Gesetz über konjunkturstatistische Erhebungen in bestimmten Dienstleistungsbereichen (Dienstleistungskonjunkturstatistikgesetz – DLKonjStatG)*

[3] *Gesetz über die Verwendung von Verwaltungsdaten für Zwecke der Wirtschaftsstatistiken (Verwaltungsdatenverwendungsgesetz – VwDVG)*

The multiple-source mixed mode method was introduced in order to reduce the response burden. The latter was only possible for small and medium-sized enterprises. The survey among the large enterprises and enterprise groups ensures the quality of the results and keeps revisions at bay.

Minor drawbacks of the administrative data, which does not fully meet the demands of short-term statistics, concern the definition of turnover in the tax prepayment notice. The tax definition of turnover differs in some respects from the statistical definition. The statistical definition does not include, for example, some extraordinary receipts (such as rental income from company-owned machinery or dwellings, sales of land or of used machines etc.). However, these are included in the turnover tax prepayment notice under the same heading as the statistically relevant services. Another – more important – issue is the character of an early estimate of the value added tax prepayment notice that depends on the unit's financial situation. In addition, in tax legislation, several enterprises can be combined in an integrated turnover tax group. In this case the breakdown of turnover is difficult. The reason is that only the controlling unit reports the total group turnover to the fiscal agency, without providing any information about how this total turnover is distributed among the different enterprises in the turnover tax group. Another issue is that the activity code in turnover tax data does not entirely meet statistical requirements. This is less a problem of standardisation – as the classifications are very similar –, but an allocation problem. A methodological test showed that the NACE code allocated to a single unit might vary depending on the data source.

In German short-term statistics on the "other services" sector the problems mentioned are addressed with regard to their effect on the quality of the results. 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. For instance: in order to produce short-term statistics in 2016 less than 1.0% of the entire enterprise population in the relevant service branches had to be surveyed directly (instead of approximately 7.5% of the enterprises that had to be questioned quarterly before the implementation). 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.

3. Quality control

3.1. Quality control of primary data

Primary data collection which provides 40-60% of turnover data serves as a sort of safeguard for the "Mixmodell". This data must be complete, accurate and timely. With regard to quality control the focus therefore lies on traditional measures. Thorough quality control of the input of every single of the largest enterprises needs to be ensured. Each unit passes through a process of automatic and manual plausibility checking. Reporting units are contacted if necessary and estimates are provided for missing data. Finally, the results are subject to macro-plausibility checks, which – when indicated – lead to a review of the input data sets.

3.2. Quality control of administrative data

Where primary data is not available, administrative data is used, virtually adding up to a full count of the enterprise population. General problems of the administrative data as described above are: incompleteness, data errors, misclassifications, non-matched data sets, estimation character of the tax data and differences in the turnover definition. The main problem here – which is not likely to be solved in the near future – is that the input data quality cannot be controlled by the statistical offices. Only in exceptional cases of major divergence the enterprises can be contacted for clarification.

Therefore, the aim of quality control in the "Mixmodell" concerning the administrative data is to produce a data set that is timely, paired with the previous period, matched with the business register data and free from significant outliers.

For this reason, the focus of administrative data quality control lies on technical control, automatic plausibility checks and estimation as well as output control. In detail, this means that the data is automatically checked for technical correctness, completeness and plausibility. Where necessary, data is automatically estimated or corrected. Correction is mainly done using the business register. It is used to match the administrative data with the data from the primary survey, recode the economic activity and distribute tax-group turnover etc. The final step includes both macro-plausibility checks of the administrative results and a manual treatment of significant outliers.

3.3. Process control and quality control of final results

The impact of the primary data and the methodological design of the "Mixmodell" compensate for deficits in the administrative data so that a high quality output can be produced. The output is only produced in form of indices and change rates (not in terms of absolute figures). As it needs to be complete, plausible, timely, coherent/comparable and accessible, monitoring the processes and output control are particularly important.

The matching procedures need to be monitored with special regard to the elimination of duplicates in survey and administrative data sources. Unmatched units are a regular phenomenon in the tax data (due to changes in tax identification numbers, new activities, legal changes etc.). However, both the rate of unmatched units and the share of turnover covered by survey data are monitored.

Timeliness is controlled during each step of the process. This is important because the timeframe is very short for the mixing process and the final checking of the indices as well as their seasonal adjustment. Even though time is short, reviews and corrections are made where necessary and possible. Primary data, administrative data and mixed data are subject to two regular revisions per quarter. The indices are also evaluated by comparison with other time-series, e.g. the economic climate indices.

4. Conclusions

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 the data

To sum up, quality control of primary data and processes can be described and ascertained by means of well-established, well-tried methods and indicators. However, the 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 the 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 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.

5. Further information

General information: *https://www.destatis.de/EN/Homepage.html*

Data: *https://www-genesis.destatis.de/genesis/online*

Further reading:

- Fischer, Hanna/Oertel, Jutta (2009): Konjunkturindikatoren im Dienstleistungsbereich: Das Mixmodell in der Praxis, in Wirtschaft und Statistik, Wiesbaden, WiSta 03/2009, S. 232 ff.

- Lorenz, Robin (2010): The integrated system of editing administrative data for STS in Germany, ESSnet Administrative data – Seminar in Rome

- ESSNET – Use of administrative data and accounts data in business statistics (2012): WP6 Quality indicators when using administrative data in statistical outputs – list of quality indicators

- Oertel, Jutta: Turnover and output measurement for "organisation of conventions and trade shows" in Germany; 30th Meeting of the Voorburg Group on Service Statistics (UN City Group) 2015, Sydney

- Oertel, Jutta: Using alternative data for German turnover indicators in the service sector; 31st Meeting of the Voorburg Group on Service Statistics (UN City Group) 2016, Zagreb