Jens Malmros
Ingegerd Jansson
SCB

Draft minutes Scientific advisory board 16-17 November 2023

Attending board members
Jan Bjørnstad
Sune Karlsson
Xavier de Luna
Steve Heeringa
Annette Jäckle
Barteld Braaksma
Per Johansson
Johanna Laiho-Kauranne
Anders Holmberg

Attending Statistics Sweden staff
Joakim Stymne
Marie Haldorson
Anna-Maria Kling
Lilli Japec
Ingegerd Jansson
Jens Malmros
Jenny Hjort
Mats Bergdahl-Kercoff
Kristina Strandberg
Can Tongur
Maria Adolfsson
Katja Olofsson
Oskar Nilsson
Petter Ehn Wingårdh
Magnus Nyström
Annika Johansson
Helén Marklund
Gunilla Sandberg
Jakob Engdahl

Other participants
Jacob Enfors, Örebro university
Sofie Orrling, Transport analysis

Krister Sandberg, Transport analysis
Maria Melkersson, Transport analysis

Welcome
Joakim Stymne welcomed everyone to Örebro and Stat Sweden´s new office. The building is new and built exclusively for Stat Sweden. It is called Kexfabriken (the Cookie factory), referring to the bakery that used to be located here.

Current issues at Statistics Sweden
SCB has appointed a new deputy director general. This position is dedicated and does not have other responsibilities. Magnus Sjöström, who previously was head of the department of social statistics, is the new deputy director general.

SCB has worked on increasing the media presence during recent years. Previously, we have supplied knowledge, but typically only as a source. There is a large expertise in the organisation, but it was previously anonymous to a large extent. The target was to double the number of media references where spokespersons were quoted. The target was reached after only one year and currently, the increase is six-fold.

Cooperation between NSIs and universities
Marie Haldorson, Stat Sweden, introduced the topic and also welcomed Jacob Enfors from Örebro university (ORU) to the meeting. There has been cooperation between SCB and ORU for long time. Recently, the scope has widened and encompasses statistics, macro-economics, and AI.

Discussion:
Are other governmental organisations part of this cooperation? This could be beneficial for other data providers because of improvements to data quality and accessibility.
-Not currently, but we could consider such projects.

Are there similar collaborations in Stockholm?
-There is no formalised collaboration but there is a PhD student in statistics.

Anders Holmberg, ABS, presented how ABS is cooperating with universities at different levels.
Anders presented the motives and the background to the ABS cooperation with universities. The benefits of university collaboration include for example quality assurance, staff upskilling and possible recruitment, and increased resource efficiency. At ABS, there is both external and internal collaborations directed at ABS staff and academics, which are funded both by ABS and research grants. External

collaborations include PhD scholarships, industry partnerships, and stipends.

Discussion
Some points made during the discussion:

- Universities don't teach survey sampling – has NSIs been unsuccessful in affecting the education?
- Universities don't teach missing data, which are always present. This is a good example of where universities fail to meet demand.
- It is important to find the right topics, where the output is useful for ABS. PhD subjects that are useful for ABS are prepared, for example, non-probability samples were studied before such methods were to be used in production.

Barteld Braaksma, CBS, presented cooperation between CBS and universities.

At CBS, cooperating with academia is a strategic priority. The cooperation is organised by an expertise centre for strategic relationships with higher education institutions. There are 12 professors by special appointment, which are recruited from within the organisation, and around 20 PhD students. In addition, there are many interns and other collaborations. There are framework agreements in place for many institutions of which the professors by special appointment are affiliated.

Discussion
Some points made during the discussion:

- If it is relevant for their work, the professors at CBS have access to data. This is a feature that is important and can help to attract people from academia.
- Funding for professors is full time, they spend one day per week at the university. This model has some effects on for example grant applications. PhD-students can be financed with different models, professors can get external financing for their projects.
- Hiring a professor is not a big change to the CBS budget, the people did similar work at CBS before they were professors.
- The professors need to have time to do research at CBS before appointment to qualify for position.
- Professors are recruited from within CBS and have existing collaborations but may contribute with outside perspectives from universities. The professors must balance between internal development and academic freedom.
- CBS seems to have a research environment within the organisation. This is not obviously easily combined with being a government agency and produce official statistics.

Topic 1: Micro Data Exchange in Foreign Trade in Goods Statistics: Potential for Redesign?

Can Tongur, Stat Sweden, introduced the topic. Sune Karlsson, Örebro University, introduced the discussion.

Discussion

Some points made during the discussion:

- The reasons to implement a new data source are that it may be better, cheaper or reduce respondent burden.
- It may be possible to use a model to speed up the monthly delivery by using VAT data. Ideally, shadow calculations should be performed for three years, which provides advice to figure how it behaves over time, e.g., on seasonal variation.
- There are three sources: VAT data which have pretty good quality with no non-response; MDE data with unreliable and varying quality; survey data for which we can control quality. Stat Sweden should figure out a way to combine these three sources. It should lead to a harmonised estimate of imports broken down by commodity etc. Break down the statistics in groups and see where the data is reliable. Use the time dimension and look at how good the data is over time.
- If the cut-off continues to deteriorate the quality of estimates this also must be addressed over time.
- At CBS, a similar discussion on user needs is ongoing. Eurostat wants data on the finest level. What do they use the data for? They publish data on the finest level of detail. Could NSIs put pressure on Eurostat on the usage of these data? Some users may need statistics on a very detailed level, but does that justify the cost of producing these data? Are we able to say that we trust the quality of these data?
- It was suggested that other areas of statistics could be compared with this case. For example, migration, flows of immigrations and emigration should match. In such statistics, matrices on how sources match (i.e., match, do not match, missing data, etc) are created and there are statistical approaches to cleaning such matrices. So far, foreign trade in goods has not found any comparable other situations.
- Another suggestion is to do a small survey of businesses, which fall below the threshold value. This has also been suggested by the ASPIRE team. There is a possibility to do a survey under the threshold, the regulation only states that the agency needs to cover 90% of trade, it does not say if that should be above or below the threshold. It is thus possible to take a statistical approach.
- Finally, it was noted that the very relevant societal question of supply chains is an illustration of the issues lifted here. It

should be the aim for the statistics to answer policy questions like that, not to publish statistics at 8-digit level.

Update and reply to recommendations on Topic 3 SAB 14 dec 2022: Need for labour – establishing a new statistical product

Martin Axelson, Stat Sweden, gave an update on Need for labour, a topic at the 2022 December meeting. This presentation replaces a written reply to the recommendations form the Board. The survey now has a new name: Job openings and recruitment needs.

Discussion

Some points made during the discussion:

- Are the recruitment plans for the next 12 months a mandatory question? It may be very difficult for businesses to answer.
  -In a previous survey, this question was posed for staffing companies and for one- and two-year periods. User feedback indicated that it was difficult to answer these questions. From the user feedback and together with the public employment agency we decided on including one easy question with the aim to simplify the questionnaire.
- Stat Sweden can take a leading role and extend this to other agencies. User needs on labour statistics are constantly evolving, for example, the European commission has proposed new rules on immigration. How flexible is the survey system in adopting new rules?
  -We are likely to be much better prepared in the new survey. For example, the third module on occupation has a flexible question to include other occupations than the standard occupations surveyed.

Topic 2: Imputation of driving distances in register data

Sofie Orrling, Trafikanalys, and Petter Ehn Wingårdh, Stat Sweden, introduced the topic. Jan Bjørnstad, SSB and Oslo University, introduced the discussion.

Discussion

Some points from the discussion:

- The first goal is to produce statistics for several domains. That is an estimation or prediction problem, rather than an imputation problem. The data are not a probability sample, so standard errors can be estimates based on a model. Multiple

imputation is a solution, but simulations at SSB showed that 'Rubin's formula' is not enough.

- The second goal is to have a complete a register. There are no missing values – the values are unavailable. If imputed values are used, is it then still a register? The proportion of missing values is high, and the variability uncertainty will be underestimated using imputation. If you regard it as an estimation problem, calculating uncertainty should not be difficult.
- Machine learning cannot be used if you have a situation where data are not missing at random. In this case, missingness is likely to depend on the driving distance.
- The first goal should rather be to have data where the quality is good enough for solving policy issues, which might mean that all data are needed. Thus, the new cars cannot be disregarded.
- Could data from other types of cars be useful? Buses and trucks have similar problems, i.e., they are not checked during the first years. Taxis on the other hand are inspected every year. Rental cars or company cars used by private persons do not give accurate data.
- Other sources could be car insurance data, data registered by the cars in real time, data collected when cars get service, or data directly from car manufacturers.
- It is possible to discuss this issue further with colleagues at CBS, where imputation methods are used.
- The main objective is to produce official statistics and not to produce registers with high quality. This would be good from a societal point of view but must be decided by government.

The process for machine learning at Statistics Sweden and the open platform BALSAM

Jens Malmros and Jakob Engdahl, Stat Sweden, presented ongoing work.

Discussion

Some points made during the discussion:

- What are the problems you are trying to solve with AI and this work? Stat Sweden is starting to frame the AI issues and will step by step enable the organisation to work with AI.
- The ethics of data and models was discussed. Algorithms can be unfair and hard to explain. This is a communication problem and not exclusively for AI but highlighted with AI. There are examples from Australia and the Netherlands where AI has not worked well. The organization need to be clear and careful with its AI policy.

- Can the continuity of results be guaranteed? There is always a risk of concept drift, models can get old quickly. But this is also not a new problem. Access to data can also change, for example if there are new costs. It might cause a total reiteration of the model.
- There is no mentioning of the AI Act in the process, but SCB is doing preparatory work for its introduction.
- At ABS there is a separation principle – separate development and production. In BALSAM, technology is used to isolate projects. They are run in a production environment to have access to data but must not mess things up.
- One of the aims is collaboration, with a federated solution, which will facilitate sharing of knowledge, models, and results. This is still in an early stage. Community standard for sharing

Concluding words

Joakim thanked the participants from the Board for their valuable contributions during the meeting concluded by giving a brief overview of the Swedish Statistical system. For more information, see the following inks:

Official Statistics of Sweden (scb.se)

Government agencies responsible for official statistics (scb.se)

The Council for Official Statistics (scb.se)

Next meetings 2024

The meetings for 2024 are schedule on May 16-17 on site in Solna, and online on November 14-15.