Statistiska centralbyrån
Statistics Sweden

Notes
Apr 20-21 2017

Advisory Scientific Board
Suad Elezović, PMU/MFS
Tiina Orusild, PMU/MIS

Meeting with the Advisory Scientific Board of Statistics Sweden April 20-21, 2017

Board members

Helen Stoye, Statistics Sweden, chair
Folke Carlsson, Statistics Sweden, co-chair
Lilli Japec, Statistics Sweden
Tiina Orusild, Statistics Sweden, secretary
Suad Elezović, Statistics Sweden, secretary
Professor Jan Björnstad, Statistics Norway & University of Oslo
Professor Sune Karlsson, Örebro University
Professor Xavier de Luna, Umeå University
Professor Daniel Thorburn, Stockholm University
Professor Thomas Laitila, Statistics Sweden & Örebro University
Professor Natalie Shlomo, University of Manchester
Professor Geert Loosveldt, University of Leuven
Professor Annette Jäckle, University of Essex

Other attendees

Professor Danny Pfeffermann, Head of the Central Bureau of Statistics in Israel (main speaker, open session on April 20)
Martin Axelson, Statistics Sweden, moderator open session (April 20)
Gustaf Strandell, Statistics Sweden
Joakim Malmdin, Statistics Sweden
Marianne Ängsved, Statistics Swedden
Cecilia Hertzman, Statistics Sweden
Eva Elvers, Statistics Sweden
Olivia Ståhl, Statistics Sweden
Stefan Berg, Statistics Sweden
Annika Lindblom, Statistics Sweden
Can Tongur, Statistics Sweden
Anders Norberg, Statistics Sweden
Magnus Ohlsson, Statistics Sweden
Nadia von Unge, Statistics Sweden
Linnea Karlsson, Statistics Sweden

Open session (April 20): Representatives from Swedish public authorities (Socialstyrelsen, Sollentuna kommun, Stockholm University, Universitetskanslerämbetet, Upphandlingsmyndigheten, Uppsala kommun, Örebro University)

Day 1

Welcome

1. Session open to all staff at Statistics Sweden and some other public authorities
    1. **Main talk**: Professor Danny Pfeffermann discussed *methodological issues and challenges in production of official statistics*.
        i. Professor Pfeffermann discussed some the major challenges in the production of official statistics in Israel. Among these challenges he mentioned the construction and the maintenance of an integrated national statistical system; permitting access to micro data under strict privacy and confidentiality restriction; reducing the response burden in business surveys; management of modern censuses; computation of house-price and house-sale indexes; making god use of "big data"; how to deal with mode effects and possible use of web panels and similar topics.
    2. Members of the Board shortly presented their own opinion about the most urgent challenges for production of official statistics in a five years perspective
2. Closed session for the Board members
    1. Discussion and advice to Statistics Sweden

3. The regular Board members meeting: Discussion about the topic of the day

Report from the discussion

The Scientific Advisory Board would like to thank Statistics Sweden for hosting the important seminar on methodological issues and future challenges in the production of official statistics starting with the invitation of the distinguished speaker, Danny Pfeffermann, National Statistician and Director General of the Israel Central Bureau of Statistics and Professor of the University of Southampton and Hebrew University (Emeritus). Professor Pfeffermann's unique insight comes from his many years of research in the field of survey methodology and culminating in his direct experiences from leading a National Statistical Institute. The topics that he covered included the following: potential uses of Big Data in official statistics production; the need for survey sampling in the future to obtain high quality data not covered in alternative sources; mixed mode and proxy data collection; issues around privacy, confidentiality and dissemination; compensating for bias in informative missing data and web surveys; university courses and training needs in official statistics.

Following Prof. Pfeffermann's presentation, Prof. Thomas Laitila from Örebro Universitet discussed specific challenges faced by Statistics Sweden which included the need for flexibility to meet demands and develop new types of statistics for incorporating alternative data sources in statistical production systems. He also focused on future statistical and data science competencies that will be needed in the workforce.

The Scientific Advisory Board members also presented their viewpoints which covered many of the previous topics as well as other challenges including the following: the growing need for methodological interventions in data collection; the need for decreasing costs in data collection and collecting data through alternative sources; timeliness and the need to publish quicker or early-use statistics; the need to research and develop multi-source and model-based estimation; develop and disseminate interval estimates for administrative and combined data sources; engagement and communication with users and stakeholders at different levels to understand future needs of official statistics.

Following the seminar and discussion, the Scientific Advisory Board of Sweden recommends that Statistics Sweden consider the following issues:

Administrative, Alternative Sources of Data and Big Data

1. Study a more integrated approach for mixed mode data collection and multiple source estimation with a guide to good practices based on available skills and capabilities within Statistics Sweden. For example, investigate the quality of tax registers as a proxy for income or the use of administrative sources to replace data collection for small businesses.

2. Develop flexible and generalized procedures that are able to adapt to new sources of data and combinations of data sources to allow different items to be estimated which follow the same principles.

3. Investigate the use of Big Data as a secondary source of data to improve prediction power in model-based and small area estimates, or other forms of estimation under combined data sources. This will compensate for the growing challenges of informative missing data and venturing into non-random sampling.

4. Given the wealth of registers and administrative sources available to Statistics Sweden, it is important to use these sources to their full capacity rather than rely on web panels and non-random samples.

Engage with users and stakeholders

5. Be more pro-active about engaging with users to investigate what are their needs and how can they be met, types of statistics and when they should be disseminated. Many of the demands are from Eurostat, OECD, etc. but a further investigation is required on how these and other demands can be met.

6. More care is needed when providing early publication of results and explaining the differences with final official statistics.

Communication

7. Statistics Sweden should consider appointing a dedicated authority responsible for external communication. This will ensure a responsible, consolidated and consistent manner of reporting official statistics. Any spokespersons engaged in external communication should work within the framework of the dedicated authority. Provide a policy of communicating results of uncertainty beyond point estimates, for example through the reporting of confidence intervals (or standard errors, margins of error).

8. **Work Force Training** The European Masters in Official Statistics (EMOS) label and accreditation has been awarded to Örebro University. This is one example how National Statistical Institutes can engage with academia on training needs and skills required for their future workforce. Statistics Sweden could be more proactive in engaging with academia on their training needs, for example by providing teaching on topics of official statistics and data collection as well as intern placements and student projects as set out in the EMOS requirements. The Board feels that this is a good opportunity to have a deeper conversation with academia and being proactive on the development of appropriate curricula.

Day 2

Introduction

Folke Carlsson summarized current issues at Statistics Sweden:

- Helen Stoye – acting GD from 15 mars until a new GD is appointed
- Report of activities (Annual report) 2016 has been delivered
- Plans 2017 and upcoming budget issues 2018-2020
- SCB is recertified according to ISO 20252:2012
- The price development for living services in tenant cooperatives in the Consumer Price Index (CPI) was earlier estimated using a proxy measure (the development of rents in rental housing)

The most important changes in the work plan for 2017

- Coordination of the short-term economic statistics continues during 2017
- Provide users with comparable statistics
- Ease the burden on respondents

Proposal for new and changed statistics where external requirements and needs have been identified – submitted 1 March 2017 . The need for funding is pressing in the following areas:

- Time Use Survey 2020
- Wealth statistics
- Business Register
- changed form of funding
- better quality
- Indicators of innovation
- Effects of national statistics with the implementation of SIMSTAT (Single Market Statistics)
- Eco system accounts

External members present own research work

Speaker: Thomas Laitila

Summary of presentation
Thomas presented his research entitled “Confidence Images of Finite Population Parameters”.

Topic 1:

Speaker: Can Tongur

Summary of presentation
The perhaps most complicated problem in disseminating foreign trade statistics is to determine the *distribution* of the below-threshold trade over commodities and countries. The methodology behind this has traditionally been to estimate so-called distributions keys for combinations of commodity and country, based on clusters of reporters. The information available on before-hand is the industrial classification code (NACE/SNI) and the aggregated trade value for reporters below the threshold, whereas above the threshold, detailed data is collected, i.e. comprising both commodities and countries.
How can collected data above the threshold be extrapolated below the threshold in a credible way?
This problem is more intricate than may appear initially and could benefit from input from the scientific council.

Questions to the board

- Can the board make any recommendations on how to obtain distribution keys?

Is the clustering approach, as presented here or in any other form, reasonable for the purpose?

- Should a completely proportional distribution approach be used as applied today (i.e. mimic collected trade) or could the distributions stem from a limited set of commodity/country combinations?

Discussant: Sune Karlsson

Summary of presentation

- Problem: Estimating the exports and imports of "below threshold traders" by product group and country.
- Known:
  - Total exports and imports for each below threshold trader (from VAT records).
  - Exports an imports by product group and country for above threshold traders.
- Proposed solution
  - Form homogenous groups (clusters) of traders based on industry and pattern of traded goods.
  - Estimate distribution keys for each group as the proportion of trade for a given country and product combination.
  - Apply the distribution keys to the known totals for the below threshold traders.
  - Refinement of current approach.
- Not clear how the distribution keys are applied to the below threshold traders (no pattern of traded goods to match on).
- The degree of explanation measure (p 6)

$$1 - \frac{\sum_{k=1}^{97}\left(\sum_{j=1}^{c}\left(y_j^{(k)} - z_j^{(k)}\right)^2\right)}{\sum_{k=1}^{97}(y_k^0 - z_k^0)^2}$$

  might be misleading as $y_j^{(k)}$ changes with the number of clusters.
- Shouldn't we care about product group/country combinations?
- There will always be missed product group and country combinations in the below threshold trade.

Concerning questions to the board

- How to obtain distribution keys?
  - How will the keys be used?
  - Clustering on information not available for the below threshold traders probably not so useful?
  - Can other information than the trade pattern be used?
  - How would a pure JATT approach work with, say, traders between 1.5 and 3 millions in trade?
- Use a limited number of country and product group combinations?
  - Probably no.
  - There may be a need to limit sensitivity to outliers but should be other ways to do this.

Additional comments

The current study essentially focuses on one approach to estimating the "below threshold" trade. While informative, this limits the usefulness of the study and a more comprehensive evaluation of different approaches is warranted. Other countries within the European Union face the same challenges and a survey of what is done in other countries should be a useful starting point and inspiration for this work.

The current approach and the investigated refinements of the current approach only use information on the industry classification of the trading companies.

Given the amount of information available to Statistics Sweden a model based approach using more background information could be a worthwhile alternative. This could usefully be complemented with a supplementary survey of the below threshold traders. Such a survey can serve several purposes. An initial survey can be used to evaluate and validate different approaches as well as facilitate model building. In a long term perspective occasional surveys can be used to verify that the extrapolated estimates for the below threshold traders are still "on track". Alternatively, the traders who are just above the threshold should resemble the traders who are below the threshold in many ways. Basing the extrapolation on the traders just above the threshold could thus be a simple and viable alternative. This is just two possible alternatives and, depending on what is found in a survey of the work in other countries, more possibilities should be considered in a more comprehensive evaluation.

Topic 2:

Speaker: Eva Elvers

Summary of presentation
Today, Statistics Sweden runs a set of monthly and quarterly surveys on the income, production and expenditures of the Swedish business population. The surveys are, to a large extent, motivated by the estimates of quarterly GDP. It has been pointed out that the current system of surveys is in need of improvement. In particular, expenditures are not sufficiently covered on a quarterly basis, affecting the reliability of the estimates of quarterly GDP.
An ongoing project has been assigned the task of implementing the production of new estimates, by collecting and processing data mainly from administrative sources. Moreover, two related surveys, on inventories and investments, should be redesigned to supply the new statistics with appropriate information.
An important question has been raised: would costs and response burden be reduced by integrating these three surveys at least when it comes to data collection? Would, moreover, integration of the surveys affect quality positively? The number of data collections would be smaller and most likely also the number of re-contacts.

Questions to the Board

- Which are the main challenges when going from "simple" surveys to complex surveys?
- Which main criteria should be studied before making a decision on a complex (integrated) survey? What in particular should the survey design try to foresee?
- Which demands are likely to be the most important ones for common methods, tools, and working routines: user needs, some quality aspect, costs or something else?

Discussant: Daniel Thorburn

Summary of presentation

- The National Accounts (NA) are based on many survey studies of the business population; At least six of them are quarterly.
  - Turnover – Inventories – Investments - Balance (Total assets) - Capacity Utilization - Economic Cycle
  - Many of them have also monthly and yearly counterparts
- There are further studies on a yearly basis
- Statistics Sweden wants to reorganize the statistics on turnover, so that costs are included and so that value added may be computed
- This will be done by dividing the population into two parts

  - Large businesses – total survey by questionnaire
  - Small businesses – from registers on Value Added Tax
- At the same time SCB wants to seize the opportunity to coordinate the survey with two other statististical products and to modernize them
  - Inventories
  - Investments
- In the large business part, the extra surveys will
  - Still be total surveys
  - Either have one new enlarged questionnaire or two new separate questionnaires
- In the sampling part the extra surveys may
  - Have separate designs – perhaps with common background questions
  - Be merged into one complex survey, where some units get the question block on inventories, some units the block on investments and some units get both blocks
    - Here every unit will have two different positive inclusion probabilities within the same survey
- Statistics Sweden wants to get advice on how to choose between these two alternatives, in this case and more generally.
- General answer to this: It is mostly good with coordination, getting more or better information to the same cost. But coordination can be done in many ways.
- The result of coordination may be called a complex survey or several separate coordinated studies. The difference is mostly semantic.
- Here a complex survey means that all objects do not need to answer all questions in a predetermined way.
- More than ten years ago we recommended Statistics Sweden to use a similar technique for ”ULF” and SILC.
- Complex may also have other meanings e.g.
  - The questions an object get depends on responses to other questions
  - Questions could be asked at different time points .
  - A rotating scheme as in LFS. (However, Statistics Sweden does not use the good design in the estimation phase).
  - Etc.
- The best way to coordinate studies depends on many factors (see below)
- To give good advice one ought to know these factors or at least take them into account
- Statistics Sweden should use previous experience to estimate some of the important factors (metadata)
- Analysis of consumer needs
  - Who are the consumers? Other than the National Accounts. What precisions demands?
  - Will a change affect other present users? Coherence and joint use with other statistics.
  - Can another change make the statistics useful to new users.
- Time aspects
  - Will the data be available at the same time or will some statistics be delayed?
  - Will a more complex design delay the data collection and/or the analysis and reporting?
  - What about changes over time? Is the same rotating scheme suitable for all variables?
- Statistical covariance.

  - If values are strongly related (given VAT-figures) much precision can be gained by merging the studies into one more complex survey
  - If the values are unrelated, there are less statistical reasons for coordinating
- Respondent burden
  - Can the same person respond to the combined questionnaire or will the reporting be divided within the company. Turnover, costs, inventories, investments?
  - What is known about the handling within businesses? If the data are in the data files they can probably be reported easily within 10 days. But if not?
- Costs
  - In monetary terms, also for
    - Non-response handling (reminders and adjustment).
    - Uneven workload at Statistics Sweden
  - In precision terms.
    - Non-sampling errors like response errors or non-response may increase due to wrong contact persons, nonoptimal sampling time ponts or question context problems
    - Good with the planned small trial survey to see what is possible!
    - Usually gains in variance if good estimation techniques are used (but will they?).
    - Also losses are possible if the coordination means that the sampling design for one question will be far from optimal
    - For the National Accounts the sum of all parts may be the interesting variable. In that case the precision of a ”complex” survey will be smaller if the parts are positively correlated
- Flexibility
  - The more complex a design is, the more difficult it is to change it, in reaction to future changes of demands and conditions
    - E.g. what happens with the use of VAT if the percentage of VAT are changed for some groups?
    - If it turns out that the businesses in the future can report figures for inventories one month earlier than for investments – can you break up the the survey again.
  - Would it be better to coordinate (with) other surveys?
    - With other quarterly surveys
    - With same topic but another period.
    - Is this a first step.

What statistical methods to use?

- In this simple form of a complex survey each object has two probabilities to be included – one for investments and one for inventories.
- The total inclusion probability of the object is the maximum of these. This is also the inclusion probability for common questions.
- A simple estimation technique is to consider each question as its own study and use the usual HT-estimator with these probabilities.
- If VAT-data are related to investments/inventory, Statistics Sweden may use a poststratified (calibrated) HT-estimator.

- Simple to use for the National Accounts since they are not primarily interested in interactions only sums of one variable at a time, but no precision gain.

Another statistical approach

- Consider all studied objects as one sample with one unit inclusion probability, but questions not asked are viewed as missing observations.
- This is a trivial form of non-response: Missing completely at random (MCAR). Estimation can now be done using standard methods to correct for missing data e.g. Multiple Imputation.
- This will probably mean using methods new to Statistics Sweden and thus more expensive to introduce.

More generally on estimation

- This is a form of cross classified sampling. Every pair of a business and a variable has its own inclusion probability.
- Simple estimates can still be used since MCAR still holds and e.g. multiple imputation is still possible to use.
- In order to know whether the new design is good one ought to decide on what estimation technique should be used.
- How much of the advantages of the new design should be used.

Questions

- Which are the main challenges when going from ”simple” surveys to complex surveys?
- Which main criteria should be studied before making a decision on a complex (integrated) survey? What in particular should the design try to foresee?
- Which demands are most likely to be the most important ones for common methods, tools and working routines: user needs, some quality aspect costs, or something else?

All these questions are quite similar. They are three aspects of the same problem.

- It is difficult to answer them in general, since situations vary between different statistical products.
- It is difficult to answer them in this particular case since there is too little background given in the paper.

Answers to a non-asked question

- We welcome this as a part of the continuous work of Statistics Sweden to improve the usefulness of the statistics within the National Accounts.
- It is areal improvement to get ”Value added” not only turnover.
- It seems to be sensible to combine the three quarterly surveys into one complex survey with three types of questions – turnover –inventories – investments.
- The main problem seems to be investments
- The first two should be possible to get directly from the economical system within two weeks. At least for large businesses. But investments requires more personal knowledge and considerations of the contact person.

Which are the main challenges when going from “simple” surveys to complex surveys?

- I have tried to list a couple of factors above. The list is incomplete but it covers many important problems.
- The most important one is as always how to identify user needs and how to best satisfy them. Everything else is a consequence of this.
- In this particular case you have probably indirectly identified many the users and found that their needs are met in both cases. Thus another challenge is probably more important here.

Which main criteria should be studied before making a decision on a complex (integrated) survey? What in particular should the design try to foresee?

- There is no single main criteria valid in most situations
- The most important thing is to take all criteria into account and not to forget any important item. You should try to foresee the unforeseen.
- I suggest that Statistics Sweden develops two standardised item lists to be used, whenever a statistical product is rescheduled
  - One list with user aspects
  - One list with other quality aspects
- To have good standards and to follow them is always advantageous also when rescheduling products.

Discussion

After the general discussion, Folke closed the meeting by thanking everyone for participating.