Statistiska centralbyrån
Statistics Sweden

Advisory Scientific Board
Suad Elezović, U/ARK
Tiina Orusild, PCA/MIH
Annika Fröberg, U/LEDN

Notes
Oct 14-15, 2014

Meeting with the Advisory Scientific Board of Statistics Sweden Oct 14-15, 2014

Board members

Lilli Japec, Statistics Sweden, chair (instead of Stefan Lundgren)
Tiina Orusild, Statistics Sweden, secretary
Suad Elezović, Statistics Sweden, secretary
Professor Frauke Kreuter, University of Maryland
Professor Jan Björnstad, Statistics Norway
Professor Sune Karlsson, Örebro University
Professor Xavier de Luna, Umeå University
Professor Thomas Laitila, Statistics Sweden and Örebro University
Professor Daniel Thornburn, Stockholm University

Other attendees

Anders Ljungberg, Statistics Sweden
Annika Fröberg, Statistics Sweden
Eva Bolin, Statistics Sweden
Folke Carlsson, Statistics Sweden
Birgitta Mannfelt, Statistics Sweden
Joakim Malmdin, Statistics Sweden
Ulf Durnell, Statistics Sweden
Ingegerd Jansson, Statistics Sweden
Karin Kraft, Statistics Sweden
Ann-Marie Flygare, Statistics Sweden
Karin Lindgren, Statistics Sweden
Jennie Bergman, Statistics Sweden
Sofia Nilson, Statistics Sweden
Frank Weideskog, Statistics Sweden
Maria Adolfsson, Statistics Sweden
Can Tongur, Statistics Sweden
Karin Andersson, Statistics Sweden

Day 1

Current issues at Statistics Sweden

Speaker: Anders Ljungberg, head of the Director General's Office

- Due to absence of Director General, Stefan Lundgren, Anders Ljungberg presented the current issues at Statistics Sweden, as listed below.
- New government announced budget cuts regarding authorities which might influence Statistics Sweden.

- Bengt Westerberg's investigation concerning researchers' access to micro data has been finalized.
- Stefan Lundgren is chair of Partnership Group 2014-2015. He is deeply involved in work with the ESS Vision.
- New Deputy Director General, Helen Stoye, starts at November 3.
- Continued problems with non-response in the surveys and costs for data collection:
  Part of the data collection for LFS will be conducted by a private company. Used mixed mode for a methodology study in the Party preference study.
    - A question from Frauke: Is it in an experimental session and what kind of sample is it?
    - Answer: It is decided that a part of the sample will be conducted by a private company (5000) and it is a random sample.
    - A question from Jan: Would this reduce the cost or…?
    - Eva Bolin's comment: In the long term some of the costs regarding the interview calls will be reduced.
- Planning for 2015: first planning period with the new strategy.
- Non-response in surveys and increasing costs for data collection are the main challenges.

Reply to recommendations

Speaker: Lilli Japec

Lilli Japec mentioned that the recommendations from the Board have been well received by Statistics Sweden.

Concerning the mixed mode in the party preference survey, some of the questions have already been discussed. Statistics Sweden agrees with the recommendations. Results from the experiment could be presented at the next meeting.
Concerning ULF/SILC, Statistics Sweden agrees that there have been some methodological flaws in the procedures. Statistics Sweden will try to implement the main recommendations from the board in the future work.

Topic 1

Disclosure: The ABS TableBuilder Protection Method at Statistics Sweden

Speakers: Ingegerd Jansson and Karin Kraft

Summary of presentation

Legal premise

- Data collected for official statistics are by the main rule confidential (Offentlighets- och sekretesslag 2009:400).
- Exceptions from the legislation make it possible to publish statistics and to use data for research purposes, if it can be guaranteed that a disclosure will not cause harm or damage to an individual, household or establishment.
- In addition: reputation of the statistical agency and trust in the system of official statistics.

Background

- Decisions on dissemination of tabular data are taken locally within the organization
- Moving towards standardization: common methodology and common IT-tools:
    - Handbook
    - Training
    - Common IT-tool mainly for magnitude data (business surveys)
    - Solution for Census 2011
- Not yet solved: common methodology and IT-tool for frequency tables
- A common solution will ensure that
    - there are no inconsistencies due to different approaches locally within Statistics Sweden
    - the staff working in production get proper support to make it possible for them to do their work
    - protection is actually being carried out where necessary, and in a manner that follows best practices.

Current work

- Frequency tables generated from the registers
    - large in size
    - large numbers of tables published on the same population
    - complex relationships between tables
    - based on totally enumerated populations
    - Necessary to have common methodology and highly automated procedures
- So far:
    - literature review
    - solution particularly suitable for Census 2011 and the publications required by Eurostat
    - a general solution: propose to focus on a methodology developed by ABS, the Australian Bureau of Statistics, and adapt it to the needs of Statistics Sweden.

ABS TableBuilder Protection Method

The protection method used by ABS has been developed in order to protect Australian census frequency data as they are made available to the public through the TableBuilder, an online system to which users submit table queries. Tables are confidentialised as they are requested by adding noise to cell values (perturbation). The same cell will always receive the same noise. In contrast to methods that randomly apply noise, the risk that somebody will be able to disclose information by repeatedly requiring the same table is thus avoided. The method was originally tailored to protect against differencing, that is the possibility to take the difference between tables for similar sub-populations and find the data for a much smaller sub-population.

Each object in the micro data is randomly assigned a permanent numeric value called a record key. When a table is requested, the record keys of the units within a cell are combined to create a cell-level key. Via a perturbation look-up table (a fixed, two-dimensional array of numeric values), the cell-level key is used to determine the amount of perturbation applied to that cell.

An attractive characteristic of the method is that it ensures that each cell is perturbed in the same way every time it is requested, that is there will be consistency between tables. A drawback is that all cells are protected separately, so that sums of protected cell values might not equal the corresponding protected

margins. Equality between margins and sums of cell values can be restored, but at the cost of consistency.

The confidentiality method was originally proposed in Fraser and Wooton (2005). The methodology is further described by Marley and Leaver (2011) and Thompson, Broadfoot and Elazar (2013). Leaver (2009) describes the implementation of the method in the Census TableBuilder.

Implementation at Statistics Sweden
Statistics Sweden has decided to start a project where the method is to be further investigated and evaluated. The technical issues of implementation are also covered by the project. The method will be tested on selected tables, covering relevant registers and table designs.

In order to design the look-up tables, Statistics Sweden need to define sensitive cells. Small values are generally thought of as being particularly sensitive and at risk for disclosure. A single object in a cell might be identified, and new information might be disclosed if the identified object can also be linked to attribute variables. Two objects in a cell can be a risk since one of them might identify the other. Larger cell values might require protection if a variable is particularly sensitive, or if all or almost all objects belong to the same category (group disclosure).

Our view is that small values (at least 1's and 2's) should always be protected. These small numbers are not what statistics is about and the 1's and 2's can have no relevance for any user. Further, even with a register of the total population, there is always some uncertainty. There has been a discussion within the agency if small values need protection or not, but if the first issue is the quality if the figures, we might as well not publish them and this discussion becomes obsolete.

The practical issues of implementing the method will need a lot of attention. The program needs to be incorporated in the production process so that the method is easy to apply, once the constraints have been formulated and a proper look-up table is defined. ABS uses SAS for implementation in the TableBuilder, and Statistics Sweden will also use SAS, at least to begin with. An interface will be developed to help the staff to apply the method. Other possibilities is to use SuperCross or SQL, these options will also be investigated further.

The units producing the tables publish a lot of tables, but they are of similar types and the same tables are published in regular intervals. It is likely that a few look-up tables will be sufficient for a large number of tables, which simplifies the implementation. The look-up tables have to be reconsidered regularly in order to catch changing demands, for example every year.

An important aspect for the units are to decide what is the most important: consistency between tables or margins equal to sums of protected cell values. This is mainly related to what the users find most important, and how the method is described and explained to the users.

Questions for the Board

- What is the general view of the Board on the ABS protection method? Is it a good choice for Statistics Sweden, or are there other methods that we should investigate further?
- Are there any details of the method that seem questionable to the Board?
- What are relevant values for the constraints of the optimization?

- Are the measures to evaluate the risk and the utility after protection has been applied relevant? Are there other measures that Statistics Sweden could take into consideration?
- How can we inform and discuss with the users?

Discussant: Xavier de Luna

Evaluation of risk: some background questions:

- Is SCB planning to have a web TableBuilder tool?
- If not: Who can ask SCB to produce tables? Who takes decision at SCB and on which grounds?
- Relevant to understand the risk of disclosure attacks.

Evaluation of risk

- Cells with count 1 and cells with small counts
- Cells with large counts:
  - grp disclosure
  - Differencing
  - Risk for SCB?

Protection

- Cells with few counts: need to be protected, but if done, the information is useless: Why allow them to start with? Why not just forbid small counts? Alternative: Qualitative information.
- If protection of large cells also needed: here adding noise is less problematic:
  - Signal to noise ratio can be controlled; trade-off between noise strength and disclosure risk.

General comments on method

- Constraints should probably be case dependent – limiting automatic application.
- Small cells allowed: then choice of constraint c difficult, and constant variance V problematic.
- Need of protecting small and large cells separately, e.g.
  - Small cells: Forbid or just bounds.
  - Large cell: constrained entropy based method.

Questions to SAB

- General view of the board to the protection method.
- Are there aspects that are questionable?
- Important to have clear picture of table production and users.
- Do not perturb small counts; forbid them or bounds.
- Relevant values for the constraints of the optimization?
  - Probably context dependent:
    - Size of the tables for instance.
    - Distribution of the observed cells (min and max count).
- Are the measures to evaluate risk and utility relevant? Other measures?
- Risk measures (signal to noise ratios, association between true and protect table) are fine but:
  - Risk of disclosure attacks must also be evaluated w.r.t. production and users (e.g. web-based TableBuilder?)
- Communication with users: information, dialogue?

    - Important to have good information to end users: Not so much how table are protected but why there are and what are the consequences for the user.
    - Dialogue: Key users may be involved in the development of the protection system.

Other comments

- Which information in SCB tables is sensitive? Not all tables need to be protected!
- What about other actors in Sweden or in Nordic countries:
    - National Board of Health and Wellfare, e.g., should have similar concerns; they have more sensitive data.
    - Possibility for collaborations?

Discussion

- Comments from Statistics Sweden
    - Ann-Marie Flygare:
        - It would be difficult task to do as suggested (focus on big numbers). Maybe not all tables are sensitive but lot of discussions are needed.
        - Lot of tables have to be published every week – it would be very difficult to sell the idea of not publishing some parts of the data.
    - Lilli: ROS has a sub-group focusing on disclosure control.
- Sune: Are the counts really sensitive? It's probably not counts but other extra information that is sensitive.
- Ann-Marie: SCB has to make sure that there is no risk at all.
- Daniel: Not sure that the proposed procedure is safe.
- Jan: All tables are potentially sensitive. Protection is important and methodology is important. Ask main users of tables – good user survey is crucial. What do we know about the actual risk of disclosure? Keep it at simple method: 0 or 3 cell counts.
- Thomas: How large is the risk?
- Ann-Marie: SCB has the responsibility not to reveal anything!

Topic 2

Editing

Speakers: Karin Lindgren & Thomas Laitila

Part I: Experiences from Selective Editing at Statistics Sweden
The main purpose of selective editing (SE) is to reduce costs for the manual work at the national statistical institute (NSI) without losing substantially in precision in estimates. A related aspect is to reduce the response burden for enterprises. Focus lies on detecting errors that have large impact on the output statistics.

To implement the idea of SE using the score function in surveys, Statistics Sweden (SCB) started developing methods in detail and a generic IT-tool in 2007. A score function considering both suspicion and potential impact was

successfully in use for the foreign trade statistics, Jäder and Norberg (2005). The first version of the generic editing tool SELEKT was implemented in the *Wage and Salary Structures in the Private Sector* survey in 2008. This survey is complex in several aspects and it was a target to have an editing tool with capacity to fully cover the needs of such a complex survey.
The SELEKT system, developed at SCB, can be seen as the HB-method generalised in various aspects. SELEKT can

(a) handle multiple sets of domains, not only a population total;
(b) handle multiple variables;
(c) apply to multi-stage designs with primary sampling units and cluster elements;
(d) combine traditional edit rules and the SELEKT-type edits which produce suspicion depending on how far out from some common dispersion an observed test variable is located;
(e) use priority weights for variables and domains in the aggregation of local scores to global scores;
(f) aggregate local scores to either the sum of local scores, the sum of squared local sums or the maximum of local scores;
(g) use a threshold at every step of aggregation of local scores so that only the part that exceeds the threshold is aggregated.

SCB has implemented SE in eleven surveys that had large spending on micro editing.

1. Foreign Trade with Goods (Intrastat)
2. Commodity Flow Survey
3. International Trade in Services
4. Wage and Salary Structures in the Private Sector
5. STS, Wages and Salaries, Private Sector
6. STS, Employment, Private Sector
7. STS, Business Activity Indicators
8. Rents for dwellings
9. Revenues and Expenditure Survey for Multi-Dwelling Buildings
10. Energy Use in Manufacturing Industry
11. Consumer Price Index (CPI)

The experience from implementing SE with SELEKT shows that pre-work such as learning about the survey, staging tables from the production database, production of performance indicators of the existing traditional edit rules, multi-variate analysis aiming at finding homogenous groups and good anticipated values and finally to find threshold values for local and global scores, are resource demanding processes.

As early as possible in the implementation stage it is necessary to address the question "Is SELEKT appropriate for the survey?" For an efficient approach to the question, Statistics Sweden has from gained experience developed a checklist and documentation templates. The checklist contains considerations of the following aspects:

(a) Micro (production) editing must be extensive, there should be a potential for savings
(b) The key variables must be continuous
(c) The main outputs must consist of aggregates of micro data
(d) It must be possible to obtain anticipated values
(e) SELEKT may have to be integrated into the production system.

If the editing mission for the survey in question has these characteristics, then the implementation of SELEKT may be proceeded with.

Fruitful outcomes

- Much needed evaluations of the editing processes while implementing SELEKT
- Decreased error lists, by 10-60 per cent.
- Decreased output editing.
- The work practices have improved, and SELEKT is appreciated by staff.

Less desirable outcomes

- Significant implementation costs, not always accompanied by corresponding cost reductions.
- Difficulties with integration in IT-systems.
- No clear way of dealing with systematic errors and inliers.

Conclusion

We have seen; less micro editing, more efficient edit rules as a result of review of the existing edits rules, pleased editing staff that considers the work as more effective, interesting and less stressful, less late macro editing, difficulties in the integration with current production systems, high implementation costs.

Part II: Devoloping Selective Editing

Experiences from implementations of Selective Editing (SE) show that it contributes with resource savings and increased timeliness. There are also some methodological questions being raised which require further attention. One issue to consider is the validity over time of threshold and parameter values in the SE procedure when they have not been calibrated against new data. Another is effects of leaving some observations unedited, a topic raised by the Scientific Board (2009) who suggest sampling of observations with scores below the threshold for editing.

This paper contains suggestions for developments of the Selective Editing (SE) method addressing these two topics. Two alternative methods are presented. One is the probability editing approach suggested by Ilves and Laitila (2009). Another technique is modeling of measurement errors suggested by Laitila and Norberg (2014). Both methods offers control of remaining measurement errors and provide with information on the performance of the implementation of SE.

Selective editing is largely based on ad hoc methods, and there is no accepted SE theory developed. In particular it does not rest on randomization theory, or any other statistical inference foundation. Thus, estimators used for calculation of estimates on selective edited data sets can be considered as biased due to remaining, unedited measurement errors.
This problem is also pointed out by the Scientific Board (2009) who suggest a probability sample of observations with scores below the threshold of selection.

Probability Editing

Conditioning on the response set of observations, traditional sample survey methodology can be applied for inference on errors in the response set. Relaxing the conditioning, results can be generalized to population and domain estimates. This probability based approach to editing is suggested by Ilves and Laitila (2009) and the theory is further developed by Ilves (2012, 2014). In terms of the

view of Granquist and Kovar (1997) the approach is an example of selective editing as only a subset of the data set is edited.
Comment: Probability editing is a readily applicable theory and can be combined with SE by simply selecting units with scores above the threshold with probability one. It is flexible in the sense it can be designed to utilize auxiliary information if available and is still applicable if such information is missing. Also, the measurement levels of the variables studied are not decisive for the applicability of probability editing.

Laitila and Norberg (2014) gives an illustration of the method. The data set used is from a stratified SRS sample survey on establishments and their salary payments to employees. The data set is merely used for illustration of the theory and the methodology suggested and the illustration is not an evaluation of the selective editing methodology in the specific application since it would require a much more rigorous treatment.

Comments: The theory and illustration show data obtained from selective editing have potential to provide information on errors in estimates due to remaining measurement errors after SE. It can be utilized for different purposes. One is quality check of estimates after SE. Large bias estimates in relation to corresponding variance estimates suggest important measurement errors in the unedited data set.

Another purpose is validating specified "parameters" in the selective editing procedure, i.e. threshold values, and parameters in local and global score functions. One particular problem is to compare estimated biases with assumed bias levels implied by the "pseudo-bias" calculated in the calibration of selective editing parameters. Finally, a third option is to utilize the bias estimates for correction of estimates.
Presently, the modeling approach described here requires a model based inference foundation for interpretation. This is due to the deterministic selection of observations. However, adapting probability editing would provide a basis for studying the problem within a randomization theory framework.

Questions for the Scientific Board
Adapting probability editing of errors poses a number of questions in relation to the practice of editing of today. Some issues of concern are

- How to handle cases with many domains in relation to the number of edited observations?
- How to handle errors identified in macro editing or by some other means after having performed probability editing?
- If editing is based on probability editing, how should Statistics Sweden proceed when making micro data available to researchers/secondary users?

The Scientific Board has earlier suggested sampling a subset of observations that falls below the threshold of editing.

- How should Statistics Sweden proceed with this work and for what purpose(s)? Should it mainly be a method for validating the SE procedure or should it also be used for adjusting estimates?

Discussant: Daniel Thorburn
Comments on

- Selective Editing: Experiences and Development by Norberg and Norberg/Laitila

I want to start by noting that editing is an important part of all data collection. Response errors are almost never unbiased which means that all estimates will have a systematic error without editing. Editing is also costly both in terms of both money and time. It is also often rather frustrating and awkward for interviewers to call back and to tell the data providers that there must be errors in the information that they have given. This may sometimes also affect the working climate and stress. It is thus of outmost importance to study and improve.

The Scientific Board has discussed this earlier several times during the last 15 years. We have encouraged further work on selective editing and recommended both further Bayesian probability assessments and probability editing. During the last presentations we have more specifically been impressed by SELEKT.

Statistics Sweden has made many improvements since then. It is very satisfactory to see all these promising results.

There exist many different situations in the production of official statistic. One consists of administrative data and registers (e.g. Self assessments but also sampling frames). In those situation. Every individual entry must be at least approximately correct, but small systematic errors may be tolerated. It does not matter if every value in a frame is five percent too low. Another situation are statistical data and registers. Here individual entries are allowed to remain completely wrong but the (estimated) total that must be approximately correct with a known standard deviation. Large measurement errors which are unbiased do seldom cause any problem. Some cases lie in-between or are combinations of the two.

Most situations lies in-between even though the statistical theory deals with the second situation. Editing is needed in both these situations but may be done differently. SELEKT is a tool which may be used in both situations but probability editing is mainly good for the second case.

SELEKT has been tested on 11 products which are described in the material. It is summarised in the following table, where I have tried to summarise the more concrete experiences. In som cases there is very little on the result. There is nothing on statistical quality in the material except some figures on pseudo bias. There does not seem to be any description on e.g. breaks in time series or whether the precision has increased. My impression is that the main goal has been to make the editing cheaper or faster but not better from a quality point of view.

Intrastat	Higher impact, same number of changes, 40 % fewer checks
Comm flow	
Int trade of services	Same number of changes, 10 % fewer checks, better work environment
Wage and Sal, Struct	Cost saving 25% (1100 h), better work environment,
Wage and Sal, S term	
Empl, S term	Lower costs
Business Act	better work environment
Rents for dwellings	Higher quality, 35 % fewer checks, global scores on higher levels
Real estate	Fewer checks

CPI

I do not know whether editing has some effects on future reporting. One might argue that a data provided who has been contacted in a previous survey will be more careful to give correct values than someone who just answered fast and carelessly and did not meet any reaction from Statistics Sweden. In the future I won´t mention that aspect on SELEKT and probability sampling.

A short description of selective editing is the following. The expected error effect of an observed unedited value is

$$p(Y \# Z \mid X,Z) * E(Y - Z|X,Z) * 1/\pi$$

where X stands for background information in the frame, Z for the observed value and Y for the true value (or more exactly the value after recontact, which still may contain errors. In Editing the value after editing is thought of as the correct value or at least the best possible value). The factors in this expression are

suspicion times error size times sampling weight.

The first term is in SELEKT called suspicion but is measured not as a probability but by something related. We have previously recommended to use the Bayesian posterior probability assessment for the suspicion part, instead of the value, which os used today. But for practical purposes most of the gain can be captured by the present approach. The last two terms together are usually called impact, but I think that it can be practical to separate them.

The expression is in fact known before the editing and can in fact be removed without any recontact, but the unknown part is described by the variance

$$(p(y\#z \mid x) - p^2(y\#z \mid x))\, E^2(y - z|x)\, \pi^{-2} + p^2(y\#z \mid x)\, Var(y - z|x)\, \pi^{-2}$$

This is the variance and the usual recommendation in Neyman allocation is to select units for editing with probabilities proportional to the standard deviation (the square root of this), but the maximum is usually flat and the procedure suggested in SELEKT will give quite good results.

Much of the new results in the material is on probability editing. One may say that there are two different ways to regard probability editing. One way is to se it as an ordinary evaluation study. The statistics is based on the unedited data, $\Sigma(Z_i/\pi_i) / \Sigma(1/\pi_i)$. The editing consist of taking a sub-sample All units in the sample are checked and the true values on Y_i are found. Finally the systematic error – bias – is found in all domains and is reported.

The other way is to regard the editing as an ordinary survey. If all values were checked, the estimate would be $\Sigma_S (Y_i/\pi_i) / \Sigma_S (1/\pi_i)$. The object of editing is to check the units in a sample (with population size n) and to try to predict what this estimate would be based on that sub-survey. This can be done e.g. with some type of a GREG-estimator. Predict $\Sigma_S (Y_i/\pi_i)$ using e.g. a GREG estimator by

$\Sigma_{se} ((Y_i - BZ_i)/\pi_i p_i) / \Sigma_{se} (1/\pi_i p_i) + \Sigma_s (BZ_i/\pi_i)$ where where Se is the editing sample, p the corresponding inclusion probabilities and B a regression matrix and ZX a vector of auxiliaries (incl Z).

It is important to know the object of the data collection. If a total survey is done and the register should be used as part of the sampling frame next year, probability editing is not so simple as this, since the object is to remove all large individual errors regardless of their weights.

Next I will go directly to the questions and my suggested recommendations. I believe that selective editing has proven its value. But there remains many possibilities to develop the methods further. We recommend Statistics Sweden to continue with this work on editing, both with implementing SELEKT to other statistical products and with developing the method further.

1. *How to handle cases with many domains in the relation to the number of edited observations for probability sampling:*

Looking at the editing sample as an evaluation study, it is the same question as what to do if you have more strata than observations in ordinary surveys. The usual recommendation is to reduce the number of strata. Domains where the effects of editing is expected to be similar should be merged. But this recommendation is only valid for the editing part. It does not affect the domains of the ordinary study,

There exist other possible answers too e.g. synthetic estimation. Estimate the function $E(\text{effect of editing}| X,Z) = g(X,Z)$ by $g^*(Z,X)$ and replace unedited observations by $Z + g^*(Z,X)$.

2. *How to handle errors identified in macro editing or by some other means after having performed probability editing.*

There may be many answers to this. The answer also depends on the procedure used when identifying the errors. (Note that in principle also suspected values which are shown to be correct should be handled like this). I will give two examples.

In the first example the error is found by looking at only one unit. One solution is to change that value from $Z + g^*(X,Z)$ to the correct one (with the old weight) . Another solution, which is a little more cumbersome is to follow the following algorithm. Remove the unit. Estimate the total as if there were $1/\pi$ units less in that stratum/domain. Add the unit to the total with the the new value but the old weight. Both methods are unbiased (if the original one was).

In the second case we consider when errors were detected on several data units. With many units involved the macro editing is more complicated, since information from many units were used and the type of covariance structure may be important. If the reason is that the sum of a domain should be within a certain interval, the units are checked one by one until this condition holds. In that case one should use both the previous procedures. But in the second case one should remove and put back all the units which are checked. (including those where Y=Z).

3. *After probability editing – How should Statistics Sweden proceed when making micro data available to secondary users*

Remember ethical aspects. A release of edited and unedited data together may for some units imply that they are shown to be cheating or it may mean that they may be identified. I do not think tht this is ethically acceptable, but I leave the disclosure topic.

One way to solve this is to give them raw data plus the result of the editing viewed as an evaluation study. This works well for e.g. occupation coding when there are many small changes. The other extreme is when only a few large errors are detected and when one is pretty certain that no important errors remain. Then the secondary user should get corrected values. A compromise is to release all checked data, Y, and all unchecked data with a flag and the estimated bias, i.e. Z and g*(X,Z).

4. How should Statistics Sweden proceed with this work and for what purpose? Should it mainly be a method for validating the SE or should it be used for adjusting estimates.

This is two questions in one. Let me start with the second one. I can´t see that there is any big difference between the two alternatives. Statistics Sweden should always strive to giving the readers the best possible estimates. So any validation should also be published. The publication can then be done as only the final estimates or a raw estimate plus the result of an evaluation study. (This assumes that the readers can add two figures, which is not always true. Many readers do not read the foot notes or method description). The choice of method depends also on the type of statistics. The second choice is simpler when presenting standard errors since the covariance term may be forgotten.

The first question is more difficult to answer. In principle, all editing should be made with probability considerations. But resources are limited and sometimes the gains with a statistical approach do not correspond to the costs.

Selective editing has proven its value. It may thus be a better use of scarce resources to first implement SELEKT to more products and to develop non-probabilistic editing further before turning to probabilistic editing. We recommend Statistics Sweden to continue this work, both with implementing SELEKT to other statistical products and with developing the method further.

The important thing is to CONTINUE!

Discussion

- Frauke: Most of modelling was not about modeling systematic error. Could or should it be done?
- Thomas: Dependence of errors among respondents. How to model: You need a lot of metadata to get good models.
- Jan: Statistics Sweden has done a very good job!
- Sune: What do we mean by systematic and random error? Actually, what we model here are conditional expectations!
- Xavier: Editing is expensive. An alternative is sequential editing: model the error sequentially until you are satisfied with results. Could we use modeling to estimate how much we need to edit?

- Karin: Editing is indeed costly. Problem today: When we implement SELEKT we actually do not know much about units below thresholds.
- Thomas: Model measurement errors to minimize amount of editing.
- Sune: Some sort of probability will be needed. Things change over time and we do not know if the model is correct or not.
- Can Tongur: We talk about probabilities below threshold. Would it not lead to bias?
- Thomas: When it comes to modeling part, you can do a lot of things without taking the data below thresholds. If we have problem with non-probability editing then we use modeling, perhaps even in the sampling stage as an auxiliary information.

Topic 3

SIMSTAT

Speakers: Jennie Bergman, Sofia Nilsson, Frank Weideskog

The acceleration of EU integration in recent decades has resulted in new challenges to national statistical institutes. Deepening EU integration calls for harmonized methods of data collection to ensure comparability of statistics across borders.

A thriving business community complying with the requirements of national statistical institutes is a prerequisite to building and maintaining a competitive EU economy. The EU commission seeks to reduce the regulatory burden stemming from EU law to improve the entrepreneurial climate and to enable enterprises to spend less time on repetitive administration. Many Member States including Sweden have launched programmes aimed at streamlining administrative procedures for business.[1]

In light of this, Eurostat launched an ambitious programme in 2012 to reduce the administrative burden on enterprise trading within EU territory whilst ensuring a high quality of EU statistics. The programme known as SIMSTAT (Single Market Statistics) aims to reduce the administrative burden on enterprises and harmonize data collection methods across the EU by exchanging microdata on intra-EU exports of goods between Member States. Member States will be encouraged to collect data on dispatches only. A 'single flow' of data will emerge as the data on arrivals will be compiled as mirror results.

SIMSTAT needs to be seen in the context of European Integration.
This implies that Sweden cannot drive the decision-making process. Sweden can give input to the legislative and decision-making process but may have to make amends in the statistical production pattern to satisfy the opinions of the majority of Member States.

Given the complexity of the SIMSTAT project, *Statistics Sweden* is faced with many methodological and administrative challenges. *Statistics Sweden* seeks the advice of the scientific council on evaluating the consequences of SIMSTAT and

[1] Näringsdepartementet, 28.05.2010, Regelbördan för Sveriges företag fortsätter att minska, http://www.regeringen.se/sb/d/13136/a/146809

how to best manage an implementation of the system from a methodological point of view.

Questions to the board

1. What is key to a successful implementation of SIMSTAT? What risks are associated with an implementation of SIMSTAT?

2. Under SIMSTAT *Statistics Sweden* needs to reduce the response burden on enterprises. A combined approach of implementing SIMSTAT and raising the threshold for arrivals can be used. Is this a viable approach?

3. How should Statistics Sweden deal with breaks in time-series and asymmetries under SIMSTAT?

4. The appendix outlines important methodological issues associated with an introduction of SIMSTAT. What is your stand on these methodological approaches?

Implementing of SIMSTAT – methodological discussion

Methodological aspect 1:

Differences in target population, coverage and cut-off levels

A variety of factors influence the decision of the cut off level:

- Quality of the data and time series that the threshold value is based on availability of data that the threshold is based on
  - trade pattern (many very large companies lead to fewer companies in the cut-off sample),
  - methods for calculating the threshold,
  - subjective assessments in the determination of the threshold value (margins or not).

Classification of the population in five possible categories:

- **Non-PSI**
  Smaller companies under the cut off value not obliged to report.
- **Reporter**
  Providers of Intrastat information having reported their trade.
- **Unit non-respondent**
  Providers of Intrastat information that have not reported their trade.
- **Item non-respondent**
  Providers that has only reported a part of their trade.
- **Unrecognized**
  No information can be found about the company.

Question to the board (1):

The effect of the different populations according to partner member states and partner companies can seriously damage the quality in the statistics. Eurostat propose that Member States must minimize the non-response. Which aspects should be considered when determining the threshold under the described scenarios?

Methodological aspect 2:

1. **Differences in estimation methods for non-response and below threshold trade**

Step 2 – Allocation of the estimated below threshold trade value (or estimation of obliged non-response traders where historical data is missing) by partner and product on the basis of Intrastat data available for traders above the exemption threshold:

- AATT approach (Adjusted Above Threshold Trade): trade pattern of the PSIs considered globally but after having excluded certain goods and/or certain PSIs.
- JATT approach (Just Above Threshold Trade): trade pattern of the PSIs just above the threshold with different options to define the JATT reference population.
- NAC approach (NACE Activity Code): trade pattern of PSIs grouped by NACE activity code.
- NAS approach (NACE Activity code and Size class): trade pattern of PSIs grouped by NACE activity code and size class (defined on the basis of the trade value or turnover).

2. **Differences in estimation methods for non-response and below threshold trade**

The AATT factors are compiled from the Intrastat declarations collected for the current month after having excluded some specific goods and/or PSIs. The goods to be excluded are the ones which definitively cannot be traded by companies below the threshold. The PSIs to be excluded are the ones whose data would obviously distort the distribution factors.

Question to the board (2):
How should the trade excluded in the distribution keys (here regarded as share of distributed trade on country and commodity level) be excluded if we use the AATT approach with respect to a purely scientific perspective in a future SIMSTAT system?

3. **Differences in estimation methods for non-response and below threshold trade**

Allowed imputation methods for non-response at total PSI level (step 1):
- Growth factor models
- Time series models or forecasting models.
- Regression models.
- Extrapolation methods based on average trade (mean value imputations).
- Imputation with administrative data, such as VAT or VIES data.

Question to the board (3):
In order to assess the different estimates at PSI level, a set of previous months' data can be estimated with each method used and compared to real reported values. A level of maximum acceptable divergence has then to be defined to allow one method to be chosen rather than another. At present an automatically selection is made in the Swedish Intrastat estimations according to the estimated and then reported values for previous reporting periods. A small difference between the estimated values and the reported indicate better reliability in the estimation method.
What's your opinion about this approach in a future SIMSTAT system?
What's your opinion about Member States estimating the non-response using one single estimation method?

4. **Differences in estimation methods for non-response and below threshold trade**

Member States are encouraged to regularly assess the quality of the administrative data (VAT and VIES) in terms of accuracy, timeliness, and where possible, comparability with Intrastat data. The reasons for doing this are threefold; to identify the most appropriate estimation methods, to measure the extent to which administrative data can be used to allocate the estimates for missing intra-EU trade and to assess the extent to which administrative data can be used to control the quality of Intrastat data.

Question to the board (4):
The coverage of some Member States is very low. This may be due to the threshold not being fixed at an appropriate level or a high level of non-response or late response. At that point, it should be underlined that the treatment of the partial response impacts the comparability of the results between the Member States. This could risk a greater partial loss of data. For many countries, it is very unclear how the procedure/methods for partial respondents is made today, and there are risks to both underestimation and overestimation. What would you advice on how to solve this problem? What estimation methods could be used for estimating partial non- respondents according to your view in a future SIMSTAT system?

5. **Differences in estimation methods for non-response and below threshold trade**

The non-respondent lacking historical information can be treated through different approaches for estimating the total trade.

The bottom-up approach regards micro level based on the aggregation of VAT data from individual PSIs. The first stage is to identify VAT registered PSIs which are required to submit an Intrastat declaration; in other words the companies above the threshold. The second stage is to identify the PSIs which have not submitted or partially submitted their Intrastat declarations. For these companies, the value of arrivals and dispatches is taken from the VAT declaration forms.

The top-down approach is based on both VAT and Intrastat data. It starts from the global value of trade obtained from VAT and Intrastat data and is therefore a macro level approach. In this approach, the value of the total NPR trade is computed as a share of total collected trade, separately for arrivals and dispatches. This share (fixed factor) is estimated on the basis of information contained in VAT declarations.

Questions to the board (5):
Do you have any particular comments to the described methods “Bottom-up approach”, “Top-down approach and fixed factor” and “Bottom-up or Top-down”? Which of them could preferably be used in a future SIMSTAT system? Are there any other approaches that could be suggested here?

6. **Differences in trade between partner Member States - asymmetries**

Questions to the board (6):
What is your viewpoint regarding the asymmetries and their impact and what methods or approaches would you suggest for work on solving asymmetries? How do you think we should tackle problems according to breaks in time series in a future SIMSTAT system considering large asymmetries between countries?

Discussant: Jan Bjørnstad

Summary of presentation

- SIMSTAT as suggested will clearly result in lower quality for import statistics
    - Unavoidable: SIMSTAT needs to be modified.
    - Main important issue: Will probably lead to biased national trade balances and hence loss of quality for national accounts and balance of payments.
- Today's data collection for Intra-EU trade: Intrastat. All Member States report dispatches (export data) and arrivals (import data). Response burden on enterprises: 50% of total response burden.
- EU want to reduce response burden for Intra-EU trade by 50 percent.
- One suggested way is SIMSTAT: All Member States only report dispatches. Use Mirror values as arrival data. Consequence: Lots of problems.

Statistics Sweden seeks advice from the Board: Eleven questions/problems are presented:

Q1a: What is the key to a successful implementation of SIMSTAT?

- Handle the problem of asymmetries.
    - Develop a method to analyze the extent of asymmetries and revise the dispatch data to get better mirror data for imports.
    - Must be able to correct mirror data for the arrivals data in a reliable way. How can this be done without actually having the true import values? Hard to see. Must use a method based on historic data and "hope" that the difference is stable in time.
    - It seems to be unavoidable to use a *combined* approach by both 1) raising the threshold for arrivals and 2) use mirror data.
- The estimates for under threshold trade and nonresponse should be handled in the same way for all Member States.
- The confidentiality issue must be handled in such a way that the PSIs still are willing to report dispatches reliably. Otherwise, the nonresponse may increase.

Q1b: What are the risks associated with an implementation of SIMSTAT?

- Almost impossible to achieve correct import data within EU trade. Is the accuracy good enough if there is a bias of 2-5% as it seems to be?
- (Unsolvable ?) problem:
    - The estimates for under threshold trade and nonresponse should be handled in the same way for all Member States.

Also, it seems that the dispatches should be sent directly to Eurostat (via NSIs). Not possible to recontact companies. It means that necessary editing is not possible.
Confidentiality problems can mean that nonresponse will increase.

Q2: Necessary to reduce the response burden on enterprises. A combined approach of SIMSTAT and raising the threshold for arrivals can be used. Is this a viable approach?

- Means that SCB suggests that they still gather data on arrivals but less than now.
- This is a good idea and I think Statistics Sweden should argue for this approach for all Member States – in fact, it is unavoidable.

Q3a: How should Statistics Sweden deal with break in time series under SIMSTAT?

- Break in time series: Parallel data collection back in time.
- For historic data: Use only dispatches from Intrastat system as mirror data for arrivals.
- Main point: Break in time series will occur because of import statistics will be less accurate.
- Not acceptable. So break in time series will basically be avoided if the problem of asymmetries can be handled.
- Consequence: SIMSTAT as suggested should not be implemented!

Q3b: How should Statistics Sweden deal with asymmetries under SIMSTAT?

- Project in EU to analyze how national arrival data match mirror dispatch data.
  - Some MS say there are large differences.
  - Some MS say there is a good match.
  - Conclusion seems to be: Not acceptable!
- Need a method to revise dispatch data to get better mirror data for imports
  - Must gather some national arrival data, with a higher threshold value than under Intrastat.
  - The *combined* approach suggested by SCB.

If reducing response burden is the only important issue: Another alternative to SIMSTAT: MODIntrastat!

- Increase threshold values for reporting both arrivals and dispatches to NSIs.
- Will get better estimates on the imports.
- Can use VAT for under threshold enterprises,
- or: Regression method with VAT as independent . Must be estimated from enterprises where both Intrastat and VAT values are available. For example in the previous years for those enterprises that then were above thresholds.
- For the future: For the regression estimation, use enterprises above threshold values «similar» to enterprises under thresholds.
- No need for microdata exchange between trading partners and thereby avoiding the most serious confidentiality issue.

The combined approach suggested by SCB

- Raising the threshold for arrivals *and* use mirror data.
- Import statistics based on nationally collected data and exchanged microdata for other Member States.
- Confidentiality is still a problem.
- Prefer MODIntrastat

If the microdata exchange between MS is regarded as important part of SIMSTAT

- Each MS makes available their dispatches data to all the other MS.

- The system for exchange of microdata needs to be embedded in the common production system for international trade on goods statistics.
    - Requires much more comprehensive production. A big challenge.
- Confidentiality considerations will be important
    - Enterprises can request not to have their data published.
    - More confidential data can cause a problem for the microdata exchange.

Some questions on methodology

QM1: What aspects should be considered when determining threshold values for the dispatches?

- Problem: Enterprises (PSIs) under threshold values in one country could be a large import value for another country:
    - Larger uncertainty on import statistics for smaller countries.
- Each data release from a member state: Cover 100% of the trade in the reference period:
    - Include estimation for nonresponse and BTT (below threshold trade).
    - Allocation of the BTT data by partner country and product group must be estimated.
    - Must lower threshold values for dispatches in SIMSTAT.
    - In fact, need some nationally collected arrival data.

QM2: On distribution to partner country and product of BTT data and nonrespondents

- Consider trade pattern of PSIs above threshold after excluding PSIs that produce goods that BTT companies cannot possibly produce.
- Consider similar PSIs just above thresholds.
- A statistical imputation method that could be analyzed:
    - Nearest neighbour imputation (NNI).
    - Main issue: Define a proper distance measure, maybe groupwise.
- For nonrespondents:
    - Trade patterns of enterprises with similar characteristics.
    - Sort of stratified NNI. The imputation method selects a statum of enterprises.

QM3: Current imputation approach for nonrespondent PSIs in SIMSTAT

- A major problem for the first publication at t+25.
- Reported data are available a month later, t+55, for about 30-40% of the nonrespondents.
- Current approach:
    - For each PSI, 5 possible categoris of imputation methods. Some depend on the availabiblity of reported data in the previous months.
    - Two of these are not good enough as earlier studies have shown, manual imputation and average reported trade of the last year.
- Remains 3 categories of acceptable methods:
    - Forecasting methods like exponential smoothing.
    - Regression type methods with VAT data as explanatory variable
        - independent error terms,
        - autoregressive error terms.
    - VAT data.
- If reported data becomes available:

    - Based on studies from 2010: Some regression type method is best when both Intrastat data and VAT data are available from earlier months.
- When Intrastat reported data are not available: Use VAT data.
- Automatic choice of imputation method, based on how the method does for previous months
    - If several methods are applicable: Need criteria to choose one of them.
    - An automatic selection is made according to estimated and reported values for previous months, the smallest difference method is chosen.
- Studies made by Statistics Sweden show that this seems to be a good approach.
- Nonrespondent PSIs with no historical Intrastat data: The only option is to use VAT data.
- Partial nonrespondents:
    - If the missing value is taxable amount use same imputation method as described.
    - For other variables: Use previous complete Intrastate data for the company.
    - Manual imputation otherwise.

Final recommendation

- Do something with the suggested SIMSTAT.
- Find altenative ways to reduce response burden.
- Remember: Some response burden is *necessary*.
- Alternative sampling plan:
    - Large enterprises always participate.
    - Stratified sample of medium-sized enterprises, coordinated.
    - Under threshold: Small enterprises.
- This could lower response burden

Discussion

- Folke: Problems not only related to response burden but also to quality. It is interesting to hear that the discussant's comments and recommendations are in line with SCB:s way of thinking.
- Birgitta: It is meant that improving quality should in long term lead to single-flow system. It is difficult to follow Eurostat's reasoning concerning this issue:
    - It's open for delivery of micro-data but we have to de estimation.
    - We do not have other countrie's VAT- data.
    - We will probably raise thresholds.
    - Important that we have the same methods within SIMSTAT. Harmonization requirement still strong.
- Sune:
    - Why should we choose only one single imputation method? This is essentially a prediction problem. Combining different methods by averaging or weighting in some way should work best.
    - Are there any other implementations of this kind of data?
    - We do not know in some cases how much we sold to another country but we have information that we have sold something. Use this auxiliary information.
- Daniel:

    - Very good idea to utilize information from different sources but there is still a question whether this really works.
    - Some goods disappear, we export more than we import according to data. Do these numbers really have to be exactly the same?
    - What is this good for? We do a god job anyway.
    - Is SIMSAT with better precision than that we have now really necessary?
    - There will be a problem with e.g. online sales.
    - Big national companies will be more problematic.
    - The question is what is going to happen with trade in the future?
    - Must do: Lowering quality and raising threshold values for special cases – special research studies.
- Frank: Tests will be done – bilateral studies – in order to solve the problem with asymmetries.
- Birgitta: It looks like the decision about implementation is already made at Eurostat but no one really knows how to handle different problems. It seems that there is an expectation that the problems will be solved as they come up during the practical work.
- Thomas: There may be a need to get input data in different ways, not in the same way as before.
- Folke: These questions are open and will probably be discussed in more details in a few years from now.
- Birgitta: We will most likely need to allocate more resources than we do today.

Lilli closed the meeting by thanking everyone for participating.