UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Seminar on New Frontiers for Statistical Data Collection
(Geneva, Switzerland, 31 October-2 November 2012)

Topic (iv): Data collection using mixed modes and multiple sources

MIXED MODE BY IMPORTING ACCOUNTING DATA INTO STRUCTURAL BUSINESS STATISTICS WEB QUESTIONNAIRES

Contributed Paper

Prepared by Johan Erikson, Statistics Sweden

I.   Introduction

1.   In order to reduce response burden, methods to use the accounting information available in administrative systems at enterprises is an attractive prospect. But due to the nature of this information and its properties, it is perhaps not possible to replace the questionnaires entirely. Therefore, ways to combine data from the enterprises´ administrative systems and questionnaires are envisioned as a way forward.

2.   The Structural business statistics (SBS) contains mostly economic data that is normally available in accounting systems. In Sweden, there is also a chart of accounts, "BAS", that is standardised (by praxis rather than by rule), but around 95 per cent of Swedish enterprises use BAS. However, some parts of BAS are not standardised in detail, it is up to each enterprise to use these parts as they wish. Therefore it is not possible to use this information alone and skip the questionnaires. An ongoing project is exploring a way to combine accounting data with questionnaires. The idea is to make it possible to import standardised export files from accounting systems into the SBS web questionnaires, and translate the accounting information into statistical variables. After that, respondents will be able to change the answers if needed and fill in the missing parts manually. The project has been granted money by Eurostat within the framework of the MEETS project and will be finished in 2013. This presentation presents the project and its progress and results this far.

II.   Background

3.   In the European Union, and in many other countries across the world, there is a political aim to lower the administrative burden on businesses. Since filling in questionnaires is an administrative burden, which is usually called response burden, the aim also covers statistics.

4.   Statistical offices use many different strategies to lower response burden, like moving from paper questionnaires to web questionnaires that can offer help during completion, making questionnaires easier, taking out questions from surveys, running surveys less frequently and moving from using surveys to using administrative data or sharing data between government

authorities. One strategy that is often wished for from respondents to surveys is to utilise the information that is available electronically in the administrative systems at enterprises rather than to ask these enterprises to fill in this information in a questionnaire.

5.   Even if the option of utilising data from the administrative systems at businesses is attractive in theory, it is not very easy to implement it in practice. There must be a translation of some sort from the information in the systems to the definitions that the statistical survey wants. And in order for the prospect to be efficient, the information should be standardised in some manner, otherwise ach enterprise would have to create their own translation. And finally, there must be some means for the information to be transferred from the systems to the statistical office.

6.   In Sweden, for economic statistics that ask questions that are related to accounting data in businesses, the opportunities are good. There is a chart of account, BAS, that is standardised and though not mandatory a de facto standard in Sweden. About 95 per cent of all businesses use BAS as their chart of accounts. Therefore, a translation between accounting data using BAS and statistical information is possible. The survey which collects most data of this type is the Structural Business Statistics (SBS) survey. Based on this, the SBS survey would be a good survey to test this possibility.

7.   The final problem to solve would be to find some standardised way to collect BAS data from businesses and transform it into SBS data. For several years, Statistics Sweden had high hopes for XBRL (eXtensible Business Reporting language) to be this standard. Some work was done by Statistics Sweden regarding XBRL and mapping of statistical variables to the Swedish XBRL taxonomy. However, this turned out not to be a way forward in the short run because only a few businesses have implemented XBRL in their accounting systems. Experiences from other countries also show that XBRL has only been implemented at businesses when it is mandatory for reporting at least some data to government authorities. However, there was also another possibility in Sweden. There is already another common format already used to extract data from the accounting systems and transfer it to for example tax authorities, accountants et cetera. The format in question is called SIE (which stands for Standard Import Export format) and it is an open standard for transferring accounting data both within the enterprise and to external parties. Due to the fact that the SIE standard is so well spread in the software business it has, like BAS, become a de facto standard in Sweden.

8.   There are some obstacles still to be met when using BAS data through SIE in the SBS survey. There are some parts of the questionnaire where the accounting data cannot easily be translated to statistical variables, due to the fact that an account with a specific number can be used with different content for different enterprises, this is for example the case for the breakdown of turnover by type of activity, this part of BAS can be used as the enterprises wishes, within some general boundaries. There are also some items in the SBS questionnaire that actually asks for a breakdown of for example some costs in even more detail than what is registered in the BAS chart of accounts. So, importing a file from an enterprise cannot completely replace the questionnaire. Therefore, a strategy that combines a questionnaire with utilising BAS data is necessary. The strategy that Statistics Sweden decided to try in the SBS survey is therefore as follows: The enterprise will receive the link to the web questionnaire as before. However, the web questionnaire will be expanded with a functionality that allows the enterprise to import a SIE file into the questionnaire. The data in the SIE file will automatically be translated into the statistical variables, and fill the appropriate fields in the web questionnaire. After that, the respondent will go through the questionnaire and fill in those parts that could not be filled in from the SIE file data and make any necessary adjustments. In the end, the questionnaire is sent in as before. This will in fact be a hybrid between questionnaire and using business data, where the enterprise can always choose to fill in the questionnaire in the traditional way if it so wishes, but allows it to use its own data from the accounting system if they wish so. But before this solution is implemented generally in the SBS survey, a number of pre-testing and testing stages is necessary to test the feasibility of the solution. The rest of the paper describes this process.

III.   First pre-test

9.   A first pre-test was carried out in the autumn of 2010, with the purpose of seeing whether information in a SIE file could actually be translated into SBS variables, and how data received in a file would compare to data filled into a questionnaire. This test was carried out mostly in-house. A number of enterprises that had already filled in the SBS questionnaire were re-contacted and asked if they could and would provide a SIE file to Statistics Sweden for evaluation purposes only. All the other work was done within the office.

10.   The pre-test resulted in some interesting conclusions:

(a)   It was harder than expected to get businesses to send in a SIE-file, and a lot of businesses (or, more precisely, the contact person in the SBS survey) did not even know what they were. This resulted in only a limited number of files actually being sent to Statistics Sweden.

(b)   The mapping of variables in the SBS survey to the SIE file was also more difficult than expected, due partly to the fact that the chart of account is changed over time and that several mappings for different versions have to be created and maintained. Some enterprises followed older versions of BAS than what we had expected, and at a later stage even those versions need to be supported.

(c)   For those variables where a mapping was possible, the data quality from SIE data was at least as good as the data provided in the questionnaire.

11.   From these conclusions, it was decided to proceed to a second stage of testing, where the whole process of using SIE files in the actual data collection was tested.

III.   Second (ongoing) test and pilot survey

12.   In August 2011 Statistics Sweden applied for, and was awarded, a Eurostat grant in order to plan and carry out a pilot study to evaluate the possibility for enterprises to submit SBS data to Statistics Sweden via file transfer. This gave us the opportunity to carry on with the project as planned. The project has four phases:

(a)   Preparations.

(b)   Pre-test – a select number of enterprises will be asked to test the solution, and cognitive interviews will be carried out with them to learn from their experiences.

(c)   Pilot survey.

(d)   Evaluation.

13.   At the time of writing this paper, only phase a is complete. Phase b is almost finished and phase c is in the planning stage, the survey will be sent out late October.

A.   Preparations

14.   In the first preparatory stage, the following activities were carried out:

(a)   Improved mapping of variables in SBS to the Swedish Chart of Accounts (the BAS Accounting Plan).

(b)   Adjustments of the web data collection tool (SIV) used by Statistics Sweden so that it can import data files in SIE format from the enterprises and upload them into the SBS questionnaire.

(c) Development of a communication plan with respondents in order to deal with the problems of communicating what a SIE file is, where it should be found and how it can be used in the questionnaire.

15. The mapping work regarding variables in the ongoing project was based on the work done in 2010 where we mapped the variables to BAS2010. However, the ambition this time is higher than in 2010, allowing several versions of BAS to be used. This means different mappings for different versions, and how those changes over time should be treated. This work is not complete before the pilot survey but needs to be continued before allowing all enterprise the SIE import alternative. Another thing that was discussed a lot was the treatment of those variables in SBS that asked for a part of an account. Other problems that we have dealt with is that the data files sometimes contains minus signs for revenues and that sometimes the SIE file doesn't state which version of BAS they are based on.

16. The SIV tool, developed by Statistics Sweden, already contains functionality to receive data files (this is used in some other surveys where respondents are allowed to send in files in a fixed format instead of filling in questionnaires). However in this project we have done some adjustments to this tool since it is not sufficient, for the aim of this project, to only receive data files. As said above, the idea is that enterprises will be able to upload a data file in the electronic SBS questionnaire using the SIV tool. In the questionnaire, the information from the SIE file is extracted and filled into the appropriate fields in the questionnaire. The respondents will be asked to complete the areas that are not found in the SIE file (i.e. breakdown of net turnover). The adjustments needed in the SIV tool have now been implemented so that it is possible to use the SBS questionnaire in the desired manner.

17. The SBS questionnaire and its design in the web tool is in itself complicated (for example, it contains a lot of calculations and edit checks, it includes pre-printing of variables and the content of the questionnaire differs between enterprises in different activity codes). Therefore, it was decided not to create a completely different version of the questionnaire for the test, something that could otherwise have been a good idea. For example, it would probably be good to re-structure the questionnaire so that those items that can be mapped to a SIE file come first in the questionnaire, and those that have to be filled in manually come later. As it is now, they came in the same order as in the ordinary questionnaire, meaning the pre-filled fields and the others are mixed.

B. Pre-test

18. The pre-test is carried out by the cognitive laboratory at Statistics Sweden and uses a pre-defined set of questions to evaluate the experiences of the enterprises. The pre-test is set up as a regular data collection, with around 10 enterprises being asked to fill in the SBS questionnaire using the SIE file import function if they wish. All material, i.e. cover letters and questionnaires, are designed exactly as they will be used in the pilot survey in the unlikely event that no need for improvements is recognised. The questions asked to the enterprises cover the following topics:

(a) SIE files in general and the respondent´s experiences of SIE files.

(b) The information given in the electronic questionnaire about the file import option.

(c) The usability of the web questionnaire concerning the import functionality.

(d) The information given and how it was interpreted.

(e) The experiences of filling in the remaining parts of the questionnaire.

19. At the time of writing, there are no general conclusions from the pre-test, but there are notes from the interviews. From those notes, it looks as if the questionnaire and the import function has worked as intended, that it was relatively easy to import the file, but that there is some

disappointment regarding to which extent the questionnaire could be filled in from the file. Some respondents clearly had expected all or almost all the questionnaire to be filled in automatically, which is not the case due to the fact that the division of turnover by activity that cannot be filled in from the SIE file is a rather large part of the questionnaire. Some information could also be clearer to be more helpful. But several test persons say that this is an improvement that makes the data provision easier. So the pre-test seems to have gone well, and the pilot survey will go ahead as planned.

C.   Pilot survey

20.   A pilot study is being organised within the annual data collection of SBS. The plan is to give a sub sample of approximately 680 enterprises (out of a total of about 16000 enterprises) the option to submit SBS data via the "import SIE file" option. We hope that at least half of the responding enterprises will choose this option.

21.   The sub sample is limited to enterprises with 5 – 250 employees. The reason for this is the discouraging experiences from the test 2010, where the very small enterprises did not understand what kind of data files they were supposed to transmit. We have therefore decided not to include enterprises with less than 5 employees. The reason for not including the very large enterprises is their complexity.

22.   We will re-design the information material and possibly also the questionnaire based on the experiences from the pre-test. As the pilot survey will start early November 2012, there is not a lot of time to do this, but the most important things stressed by the test persons will of course be addressed.

D.   Evaluation and beyond the pilot survey

23.   In early 2013 the pilot study will be evaluated and decisions on how to proceed further will be made. The expected effect on the respondent costs and production costs will be estimated. A final report will be sent to Eurostat by the end in May 2013.

24.   The project will also, as the next step, provide a specification of requirements that can be used by software providers to create statistical reports directly in the accounting systems, to further increase the number of variables that can be reported more automatically.

25.   If the pilot survey goes well, offering the SIE file import option on a larger scale will probably be implemented in 2013. Before that is done however, there is some work that is already envisioned that will have to be done:

   (a) Further improvement of the mapping of variables and expansion to even more versions of the BAS chart of accounts.

   (b) Improvements of the SBS web questionnaire to adapt and optimise it for the SIE import option.