Session 3 Business Statistics

Seminar on Registers in Statistics - methodology and quality 21 - 23 May, 2007 Helsinki

A strategy to increase the use of administrative data within an integrated system of business statistics

by

Svein Gåsemyr,sga@ssb.no, Sindre Børke, sbe@ssb.no, and Morten Q Andersen, moa@ssb,no Statistics Norway

1. Introduction

The Norwegian government wants to reduce the response burden for businesses. At the moment we see a strong development of the infrastructure to promote electronic reporting. One measure to reduce response burden is to ensure coordinated reporting to government agencies. An example, if an enterprise reports a variable to agency A, other agencies are not allowed to collect data on the same variable. If another agency such as the NSI needs this variable it has to be collected from data systems of agency A.

Administrative data for the units of enterprise and establishment are used both as a source for register based economic statistics and as supplementary or auxiliary sources for business surveys. Administrative data is defined as data collected by government agencies for administrative purposes. This paper focuses on use of administrative data in processes of the statistical value chain of business surveys. Most of the processes of a traditional business survey should be improved by use of administrative data as auxiliary or supplementary sources.

The Nordic countries have been in the forefront in **i)** developing infrastructure for electronic reporting, **ii)** cooperation between government agencies in data collection and **iii)** the development of register based statistics. So far the register based statistics are most developed in population and social statistics, see [1] and [2]. At the moment there is strong development in the Norwegian register based economic statistics. A milestone in this development is the statistical database for enterprise accounts.

One trend in the Norwegian development of the statistical value chain is to increase the use of generalized survey software that is designed in modules to be used when needed. One project of this kind is the Dynarev project. Dynarev in Norwegian stands for Dynamic Editing and the first step covers 15 price statistics. Dynarev is to be used for most business surveys. The Dynarev editing and imputation procedure is linked to most of the processes in the value chain of a business survey and affect to some extent most of the processes. The examples on increased use of administrative sources in business surveys to be presented in this paper focus on annual structural statistics of the manufacturing industry.

The paper will present the interplay between **i**) base registers, (and other infrastructure) **ii**) a database to coordinate the sample design for business surveys, **iii**) a database for enterprise accounts and **iv**) a menu for data editing and imputation. The database for enterprise accounts has been developed for several years and covers now almost all Norwegian

enterprises. The paper will concentrate on the use of administrative sources in processes of business surveys.

2. Infrastructure to promote efficient enterprise reporting to government agencies

Legistation
All administrative data systems and base registers are related to some legislation, usually it is specified which variables government agencies and private sector have access to. The infrastructure of base registers and use of official ID numbers needs to be combined with an efficient and reliable legislation to protect privacy and confidentiality.

E-government
The Norwegian government has followed an active policy to develop e-Government aiming to reduce response burden and ensure efficient administration. Principles to ease the burden of reporting obligations on enterprises:

- The Government shall never request more information than is utilized
- The respondents shall never have to report the same information more than once
- The Government shall offer the most efficient means of reporting
- There shall be correlation between the use derived by the Government from the requested information and the burden it imposes on the respondents.

Electronic reporting
In order to offer the most efficient means of reporting a portal has been developed. The portal, AltInn, is to be used for electronic reporting by enterprises to government agencies. AltInn should provide a way of reporting data electronically from enterprises and government organizations. This involves statistical questionnaires, application forms, tax forms etc. A separate portal to be used when a household needs information from government sector or when reporting to a government agency, is just in operation. This portal has some relation to AltInn.

The application architecture consists of:

- The portal where data providers can fill in forms and submit data
- Integration of exchange of data between enterprise systems and administrative registers where available
- A communication service center/archive for handling reported data
- The forms and maintenance of the forms
- A common security infrastructure (PKI) for eGovernment.

Base registers and use of official ID numbers
In Norway measures to promote use of electronic systems and reporting such as base registers and use of official ID numbers for the most important units, person, enterprise and building/address started in the 1960's. The official ID numbers are used in administrative data systems in the government and private sector.

The statistical Business Register serves as frame used to design and draw samples for business surveys and administrative data estimation.

Statistical units and variables covered in administrative data systems and base registers
The statistical Business Register is more or less an integrated part of the administrative Legal Unit Register (LUR). Statistics Norway has succeeded in having the statistical unit of establishment registered in the LUR. Statistics Norway is responsible for profiling an enterprise in two or more establishments when suitable. The profiling of an enterprise is common for the BR and the LUR. The statistical unit of establishment is registered in the Social Security data system on employee jobs. The unit of employee job is linked to the statistical unit of establishment. This is an important advantage for register based employment and production statistics. As there is no need for the statistical unit of establishment in administrative procedures Statistics Norway has to use resources to watch over the reporting of complex enterprises on employee jobs. In the future there might be some administrative needs for the unit of establishment.

A Metadata Information System for government sector
All units and variables that an enterprise has to report to the government sector have to be defined and described in a common Metadata Information System for government sector. The system is seen as a tool to measure and reduce response burden and to easy electronic reporting.

The Brønnøysund Register Centre (BRC) is now developing SERES (Semantic Register for Electronic Exchange), a common system for uniform modelling of information, which will provide harmonised, standard definitions for exchange of information by the public service. This does not depend on whether the information is to be sent from the public to the government, between different government departments or between machines within the same government department. All government departments are invited to cooperate in this work to ensure that the data they are dependent upon are documented within SERES. Methods and tools are being developed in cooperation with the most important, branch specific communities, to find good approaches and solutions that take into consideration both national and international work in this area.

For correct and consistent use of data from different sources in different systems, it is extremely important that departments do not establish local definitions for data that actually have the same meaning. The ability to find and re-use definitions that are registered is dependent upon departments having good and simple access to the information models that are already established centrally (standard model) and for other departments (domain models).

3. Supporting or auxiliary sources for business surveys

Administrative sources of interest for business surveys are enterprise accounts, general trading statements, data system on VAT, employment, and wage sum (i.e. compensation for employees). Statistics on foreign trade has always been based on administrative sources. Survey data from previous periods are used to identify large changes of a former reference period for a given establishment or enterprise.

The database for register based labour market statistics is the source for employment data in all business surveys.

The statistical database of enterprise accounts
The database on enterprise accounts covers almost all enterprises in the private sector. There are two types of sources for enterprises that are liable to pay tax, **i)** annual reports for companies obliged to report to the Register of Company Accounts and **ii)** general trading statements (tax questionnaire on accounting). The general trading statement for the tax agency is more detailed than the annual account. There are several types of general trading statements depending on economic activity and whether or not the enterprise is obliged to follow the Accounting act.

The Tax Agency, the Register of Company Accounts and Statistics Norway have worked for many years to coordinate the reporting of enterprise accounts and general trading statements to the government sector. This cooperation includes electronic reporting and reporting direct from internal enterprise data systems. One result of the cooperation on accounts data is the electronic portal for government sector, AltInn. The first version of AltInn was developed to report enterprise accounts and general trading statements by electronic reporting.

For single establishment enterprises we find consistency between enterprise accounts and production variables reported for the unit of establishment. For complex enterprises there are no accounts for the unit of establishment. The staff responsible for accounts statistics states that it is impossible to distribute accounts variables of a complex enterprise on two or more establishments. Nevertheless people responsible for production statistics sometimes distribute accounts variables on two or more establishments. National Accounts needs consistency between production data, that is collected for the unit of establishment and accounts that are collected for the unit of enterprise.

An extended database of enterprise accounts
One of the ideas in the Dynarev project, see chapter 5, is to organize all available data for an enterprise and an establishment in one database. The data for an enterprise and establishment are needed as a time series and the variables have to be organized as a longitudinal database.

4. Sampling frame and coordinated sampling for business surveys

The system of 3 integrated statistical base register, person, Central Population Register, (CPR) establishment, Business Register (BR) and property, address and building (GAB) are updated by administrative base registers for these units, a common data model is created and they are organized in a common database.

The 3 base registers are integrated by using the ID for address from GAB in CPR and BR. The register based job file represents the link between the CPR and BR. The unit of job is identified by the PIN of the person, the BIN of the work place and the period for which the job was/have been active. The job file represents a very interesting integration of social and economic statistics.

A common database to coordinate sampling for business surveys have been in operation for some time. A project to improve the coordinating is in progress. The data collection strategy of Statistics Norway aims to reduce the response burden for enterprises. Measures to achieve the objective are:

- reuse of administrative sources
- promote direct reporting from enterprise data systems

use of electronic track of third part
promote electronic reporting
rotation of sample to distribute the response burden to many enterprises

The Canadian Unified Enterprise Survey should be a model on how annual structural business survey can be coordinated, see [3].

4.1 Sampling design for structural manufacturing industry statistics

During the last 4 years the sampling rate for medium size enterprises has been reduced. The reduction is based on a controlled process to be informed on how the reduced sample affects the quality of the statistics. The variables are based on **i)** the BR, **ii)** the survey **iii)** administrative data and **iv)** survey variables by mass imputation.

Table 1. Number of enterprises and establishments

	Number of enterprises		Number of establishments		Percentage of production value	
	2001	2005	2001	2005	2001	2005
Complex enterprises	1 056	1 086	1 905	1 768	55.5	59.4
Single establishment enterprises	20 052	19 416	20 052	19 416	44.5	40.6
Total	21 108	20 492	21 957	21 184		

Auxiliary units are not included in the number of establishments, but are included as units in complex enterprises in the figure percentage of production value.

Enterprise accounts refer to the unit of enterprise. For single establishment enterprises the enterprise accounts refer also to the unit of establishment.

Table 2. Sample design and estimates

	Number of establishments		Employment		Percentage of Production	
	2001	**2005**	**2001**	**2005**	**2001**	**2005**
Complex enterprises	1 905	1 768	129 853	119 107	55.5	59.4
A. Census survey Sample population	1 620	1 297	128 609	111 870	55.2	55.5
B. Selected for survey	0	97	0	4 508	0.0	3.0
C. Not selected	0	55	0	800	0.0	0.2
D. Small units use of simple form	285	319	1 244	1 929	0.3	0.7
Single establishment enterprises	20 052	19 416	157 982	140 551	44.5	40.6
A. Census survey	3 071	799	112 291	66 826	35.7	25.4

Sample population						
B. Selected for survey	0	1 154	0	21 038	0.0	4.6
C. Not selected	0	1 037	0	17 999	0.0	3.7
E. Excluded small units	16 984	16 426	45 691	34 688	8.8	6.9

All complex enterprises participate in the survey. Small complex enterprises respond to a simplified questionnaire, 285 establishments in 2001 and 319 in 2005. Other establishments of complex enterprises and large single establishment enterprises are surveyed by a census. In 2001 there were no uses of sampling. In 2005 medium-size establishments are surveyed by sampling.

Table 3. Figures by sources, Business Register, register based job file, survey and mass imputation

	Number of Establishment Business register		Employment in 1000. Register based job file		Production in per cent Diff sources	
	2001	2005	2001	2005	2001	2005
A. Census survey	4 691	2 096	240.9	178.7	90.9	80.9
B. In sample and selected	0	1 251	0	21.0	0	7.6
C. In sample and not selected	0	1 092	0	18.0	0	4.2
D. Small establishments use of a simple form	285	319	1.2	1.9	0.3	0.7
E. Small establishments excluded	16 984	16 426	45.7	34.7	8.8	6.9
Total	21 960	21 178	278.8	254.3		

The number of establishments that participated in the survey were 4 691 in 2001 and 3 347 (2096 + 1 251), in 2005. The sampling of medium-sized establishment is introduced to reduce response burden. Small single establishment enterprises, the exclusion threshold is 10 employed persons, do not participate in the survey.

Sources of variables in the statistical micro file

The number of establishments is based on the statistical Business Register. Other variables that are based on the Business Register are industry and institutional sector. Statistical surveys are used to update the administrative Legal Unit Register and the Statistical Business Register on these variables.

Employment, measured by employed persons and hours worked, is collected from the register based job file. Small single establishment enterprises do not participate in statistical surveys. This practice entails a challenge in updating the NACE code.

The survey is the source for variables collected in the survey, group A, B and D in table 3. Unit and item non-response are corrected by imputation at unit level. For simple establishment enterprises most of the survey variables are collected from the enterprise accounts or the general trading statements. Some variables for these units such as production, are based on mass imputation.

The method of mass imputation is a simple rate estimator. The imputation is based on 5 digits NACE code, (3 digits code for some industries).

How to decide the exclusion threshold

The Canadian Unified Enterprise Survey use a systematic method to decide exclusion threshold. Administrative data from enterprise accounts are used to estimate for the smallest establishments within the population, those cumulative representing the lowest percentage of total revenue, e.g. 5 percent and 10 percent of total revenue.

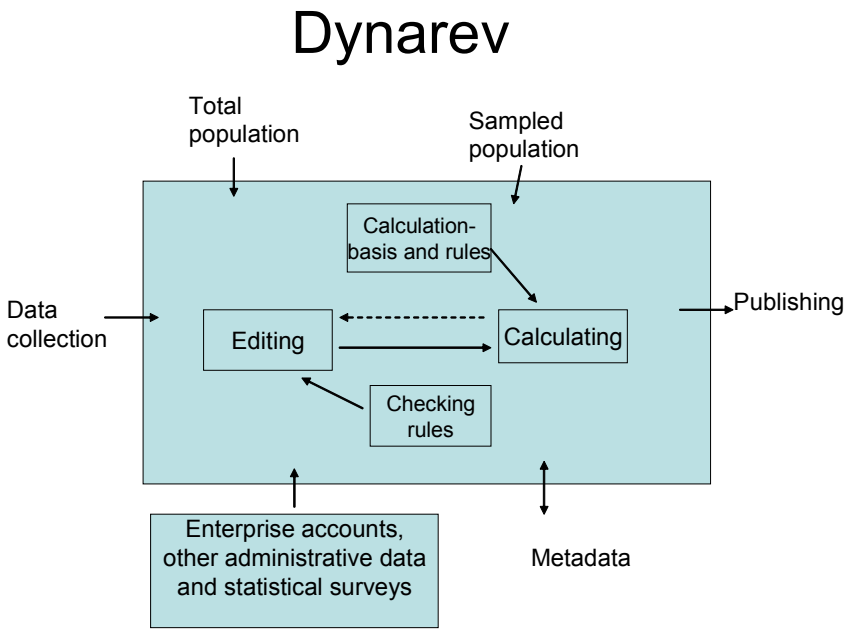
Updating on industry code for small single establishment enterprises

One problem related to the fact that small single establishment enterprises do not participate in statistical surveys is that there is no source for updating of NACE code. Information of the quality for variables such as address and NACE code for small single establishment enterprises are needed.

5. The Dynarev project

The term Dynarev in Norwegian language stand for Dynamic Editing – in English the term could be Dynaedit. The Dynarev project is developing a generic system (figure 1) that covers all processes from a filled in questionnaire is received in Statistics Norway ("data collection") to the final statistical micro file is finished ("publishing"). The system is designed as a menu of modulus that can be selected for specific statistics. The Dynarev should be seen as a tool to increase the use of administrative data in business surverys.

Figure 1



Editing using data from different sources
The Dynarev presents a flexible tool to design controls and editing for data from the specific survey. There will be an opening for complex editing and imputations, based on information from different/several sources.

Information from the base registers (population) will be available in Dynarev. A database on enterprise accounts is in operation. It is proposed to include all available administrative and survey data of the units of enterprise and establishment in this database. This will be a fundament also for editing and imputation procedures for the micro file of a business survey. Consistency between different sources and changes from former survey periods can be checked.

In Dynarev, as in most systems for editing and imputation, there will be registered some process variables. Both statistics on the editing process and identifications of changes being done are important. Statistics Norway is trying to improve the reporting of such information through a standard set of quality indicators for the data collecting process.

Calculations.
The calculation rules are constructed (tuned) from a standardized application and executed based on information from the total population, the sampled population and other sources.

Dynarev is developed for, and until now used in 15 surveys on prize indexes. An alternative calculating application for structural statistics is under construction.

Dynamic revision system.
In Dynarev, as in general, the editing procedures should include different statistical controls, focusing on critical variables and units in the survey. The identification of unusual changes from earlier data on unit level is a part of this. Editing using graphs is included in the concept. Further, the Dynarev facilitates controls on aggregated (calculated) data, identifying variables that seems to have unexpected changes, following up by inspections on underlying data. This gives an iterative process through editing and calculating.

Statistics Norway will be able to present this system in a later UNECE Work Shops on data editing and imputation.

6. Implementing business survey data collection

Several modes are in use for statistical data collection from enterprises and establishments. For all methods of enterprise reporting the key question is about the availability of variables needed for business surveys. The internal enterprise data systems usually have efficient solutions for reporting to government agencies such as Tax, Social Security and Customs. The preparedness to report to statistical surveys seems to differ a lot between enterprises. Small enterprises often have problems filling in a statistical form. Information about input and output products and services needed for European structural statistics and National Accounts are not needed for other government agencies and often entail an extra response burden. To succeed in promoting increased use of electronic reporting Statistics Norway needs better and more systematic information on the problems a enterprise is facing has when reporting to statistical surveys. One effect of electronic reporting is that part of the editing process is moved to the respondent. Some NSIs are talking about response and edit burden. Electronic reporting should include improved systems for communication between enterprises and the NSI.

Paper questionnaire
is the traditional mode for data collection in business surveys.

Electronic questionnaire
on web and e-mail is used for reporting by more and more enterprises. Some years ago the Norwegian government decided that for *all enterprise reporting* to government agencies there should be an electronic alternative by the end of 2004. Statistics Norway fulfilled this demand by 1 July 2004. Some of the resulting electronic questionnaires were not very well designed.

Direct reporting from internal enterprise data systems
through the portal AltInn is the most efficient way of reporting both for the enterprise and the government agency. A prerequisite to succeed in a strong development of direct enterprise reporting is to establish a good, efficient and continuous cooperation between the most important government agencies, the most important software companies for developing enterprise data systems and enterprise organizations. Electronic reporting of enterprise accounts is an example where this kind of cooperation has resulted in a good practice.

Other electronic reporting
A second example of cooperation to promote direct reporting is use of enterprise data systems on staff and wages in reporting to Tax, Social Security and wage statistics. Statistics Norway established (in the late -90's) a fixed-record, diskette-based reporting system (from enterprise

systems). This practice delays the introduction of more modern system solutions of today. There is a project in progress that aims to organize the reporting through AltInn.

Use of electronic tracks of third part
Statistics Norway has started to collect data from electronic tracks of third part. It is expected that this mode of data collection for statistical purposes will become more and more important, see [4].

7. Estimation by mass imputation

One advantage of estimation by mass imputation is that the resulting micro files cover the whole population, i.e. we have census files. This means that an unlimited number of consistent statistical tables based on the business survey can easily be produced.

The method of mass imputation represents a method for an optimal utilization of available sources of base registers, administrative sources and earlier survey data

The traditional way of executing a business survey is to carry out a census for large enterprises, a sample survey for medium size enterprises, an estimation of medium size enterprises and use of available data for small enterprises. The data collection described in this paper for structural manufacturing industry, could be seen as turning upside down the traditional approach for business surveys. The starting point is available sources of base registers, other administrative sources and earlier survey data. These data cover all enterprises and establishments of the population. The business survey is seen as a supplementary source for units and variables not covered by the starting database.

What should be the most efficient mix of census and sample survey? How far could we go in deciding a high exclusion threshold? Statistical methods on these topics should be more developed.

Referanser

[1] To be published by UNECE (2007): Register-based statistics in the Nordic countries – Review of the best practices with focus on population and social statistics.

[2] Eurostat, Danmarks Statistik (1995) Statistics on persons in Denmark: a Register-based Statistical System.

[3] Statistics Canada (2005) The Canadian Integrated Approach to Economic Surveys

[4] Yngve Bergstrøm, Statistics Norway (2007) Datafangst fra internett