

## *Seminar on Reengineering of Business Statistics, Lisbon 11. - 12. October 2007*

<b>SEMINAR ON REENGINEERING OF BUSINESS STATISTICS, LISBON 11. - 12. OCTOBER 2007 .....</b>	<b>1</b>
<b>FINNISH SBS SYSTEM; USE OF ADMINISTRATIVE DATA, METHODS AND PROCESS .....</b>	<b>1</b>
1 THE SURVEY OBJECTIVES, ORGANISATION AND COSTS .....	1
2 THE TIMETABLE OF SBS PROCESS AT STATISTICS FINLAND .....	2
3 DEFINING THE SURVEY TARGET POPULATION USING BUSINESS REGISTER .....	2
4 SURVEY SOURCES.....	3
4.1 Direct inquiry.....	3
4.2 Tax Authority data as administrative source .....	3
5 STATISTICS FINLAND'S DIRECT INQUIRY AND TAX AUTHORITY'S QUESTIONNAIRES .....	4
6 IMPUTATION METHODS.....	5
6.1 Classifying observations .....	6
6.2 Unit non-response.....	6
6.3 Item non-response.....	7
7 ESTABLISHMENT STATISTICS (LOCAL KAU) .....	8
8 MULTI YEARLY TRADE STATISTICS.....	8
9 PRELIMINARY DATA COMPILATION .....	8
9.1 Structural imputation coefficients.....	8
9.2 Use of the VAT files .....	8
10 OUTPUT AND DELIVERIES .....	8
10.1 Financial Statement Statistics (FSS).....	8
10.2 Regional Statistics on Manufacturing.....	8
10.3 Structural Business Statistics (SBS).....	8
11 BENEFITS AND CHALLENGES IN USING ADMINISTRATIVE DATA FOR BUSINESS STATISTICS.....	8

## *Finnish SBS system; use of administrative data, methods and process*

### *1 The survey objectives, organisation and costs*

Objective of the SBS survey at Statistics Finland is to full fill the requirements of the EU CR 58/97 annexes 1-4. Also domestic users are important, National Accounts including Regional Accounts and Financial Accounts being the most important customers of the SBS products. The Financial Statements Statistics (core of SBS) survey cannot solely full above mentioned data needs. Other data sources such as R&D data, environmental costs, energy data and external data on sales space are used to bridge the gap.

The SBS data in Finland is compiled using direct inquiry and tax data combined with Business Register's classifications. Also budgets of municipalities and joint municipalities boards are used. Direct data treatment involves around 14 persons working on manual data collection, editing and dissemination. Tax data is treated automatically using mass editing and imputation techniques. These automated routines requires 2 person working on the maintenance and development of the system.

The manual data treatment is working in the Sybase environment with the self built graphical user interface (GUI). On the other hand the automatic data editing and imputation uses the Unix server with the SAS software. In practice the automated

c:\documents and

edits and imputations consists of various SAS programs which are submitted to compile the data. The major challenge of the Finnish SBS system is to further expand the automated system limiting the manual data treatment.

## *2 The timetable of SBS process at Statistics Finland*

The timetable of the SBS data production at Statistics Finland for ref. yr. 2005 was:

- Production frame is drawn from business register (August 2005)
- First questionnaires are sent to reporting units (September 2005)
- Production frame is updated from Business Register November 2005;
- 1st download from Tax authority (16th Feb 2006)
- Last questionnaires sent to reporting units (February 2006)
- 2nd download from Tax authority (18th Apr 2006)
- 3rd download from Tax authority (15th May 2006)
- 4th download from Tax authority (24th May 2006)
- Preliminary results (27th Jun 2006)
- Preliminary data for LKAUs (28th Jun 2006)
- Preliminary SBS delivery to Eurostat (October 2007)
- 5th download from Tax authority (1st Aug 2006)
- 6th download from Tax authority (1st Nov 2006)
- Closure of the survey data treatment (November 2006)
- Business Register completed, final frame (November 2006)
- Survey, taxation and Business Register data combined (November 2006)
- Publication (15th Dec 2006)
- SBS data delivery to Eurostat (June 2007)

*Note: In FI the accounting year is not necessarily same as the calendar year. That is, the legal units deliver the data to tax authority 4 months after ending their accounting year. To synchronise SBS data production with the tax data collection Statistics Finland is ordering several data batches from the tax authority's database.*

Direct survey production base is updated from Business Register in November 2005. The data is used for the data collection, checks etc.. There are two other updates from Business Register in February and August. For 2005 there were 6 batches from Tax authority's database, the 1st download was in 16.Feb.06.

In general the questionnaires are sent to reporting units after the units have ended their accounting period. The last questionnaires for 2005 were sent in February 2006. Preliminary results were published in 27.Jun.06. The coverage was 78% of enterprises and 95% of turnover. This implies that the emphasis of data collection and treatments is on the larger units. Missing units were imputed using turnover change from value-added tax data. The changes were augmented with structural change estimates from the received data.

Final frame was drawn in November 2006 when the Business Register was finalised. Survey, tax and Business Register data was combined in November 2006 and the Financial Statements Statistics (core of the SBS) were published in 15.Dec.06. Finally the SBS data was transmitted to Eurostat in June 2007.

## *3 Defining the survey target population using Business Register*

SBS survey consist of two main sources (or sub-surveys) the administrative tax data and direct inquiry. All units included in SBS survey are drawn from the Business Register.

The preparatory target population consist of all active enterprises including following legal forms natural persons, decedent's estates, corporations subject to taxation, general partnerships, bankrupt's estates, limited partnerships, shipping companies under joint ownership, limited companies, co-operative societies, economic associations and government enterprises. That is, housing corporations are excluded from the survey target population.

Finally, the smallest enterprises are excluded from the frame. The enterprise is excluded if none of the three minimum constraints is full filled. Minimum constraints are consist of turnover, balance sheet total and number of personnel requirements.

## 4 Survey sources

The SBS in Finland is a census survey. That is the data for all units included in the frame is collected either directly or using administrative sources. This is paramount for the precision, cost efficiency and response burden of the data. Good precision guarantees better data for the entrepreneurs, micro and small enterprises. Otherwise the reliability for the estimates would be low including relatively high variation error. Also the cost efficiency can be improved and the response burden can be reduced using administrative data. Especially the response burden for the micro and small enterprises is rather low because only enterprises employing more than 20 persons are included in the direct inquiry.

### 4.1 Direct inquiry

The threshold is applied in the direct data collection. Data is collected for the units employing more than 20 persons for activities C - F (NACE 1.1). For the activities G, H, I and K all enterprises employing more than 50 persons are included and the largest units from the group employing 10 - 50 persons according to turnover size. For the activities M, N and O only the enterprises employing more than 50 persons are included.

The survey concerns the accounting period that ended during the statistical year. In the case of multiple accounting periods for one enterprise during the year, all periods should be taken into account. Also, when the accounting period is shorter or longer than 12 months and the operations of the business are as usual, the accounting period is transformed into 12 months using linear extrapolation.

The direct inquiry is collected using uniform questionnaire for all activities. The questionnaire has 120 questions concerning details on income, expenses investments and personnel. The web address of the questionnaire is included in the chapter 5.

The direct inquiry is collected and treated manually. Currently electronic data collection via internet is piloted with the inquiry. For local KAUs 57,1% (third year) and for enterprises 18,1% (first year) of the data was received in the electronic form for the ref. yr. 2005. For the ref. yr. 2006 the percentages are 75% and 36%.

Unit non-response for the ref. yr. 2005 was 15%. The data for these enterprises is estimated using tax data and imputations. In the case of item non-response first the data is asked via reminder letters (three) and then previous year data is used. If no preceding data is available usually mean imputation is applied.

### 4.2 Tax Authority data as administrative source

Six batches of the tax data is downloaded from the Tax Authority's server each year. The timing of the batches for ref. yr. 2005 see chapter 2.

Tax authority data includes information on profit and loss account, balance sheet and investments. The data is received for all the enterprises paying income taxes in Finland and that have ended their accounting period during the statistical year.

So far the tax data has included various errors. That is, the automatic data edits are needed to reach feasible data accuracy. These routines are based on the linear equations and accounting identities. First all erroneous observations are separated from the good ones. Then errors are edited using the following preference order.

- 1) Logical edits
- 2) Outlier detection
- 3) Small errors (<5%) from turnover are re-scaled
- 4) Donor based imputation
  - 4.1) Previous year
  - 4.2) Nearest neighbour

Nearest neighbour imputation implies that the distance measure needs to be developed. The measure used here is presented in the chapter 6.2.

<i>Editing quality category</i>	<i>Count</i>	<i>Turnover</i>	<i>Turnover (excl 1)</i>	<i>Turnover (excl 1 &amp; 2)</i>
1 Direct Inquiry	2.9 %	74.6 %	0.0 %	0.0 %
2 Tax data, consistent	80.8 %	20.2 %	79.4 %	0.0 %
3 Tax data, outlier detection	2.7 %	0.9 %	3.4 %	16.4 %
4 Tax data, error less than 5 percent	3.0 %	2.6 %	10.4 %	50.4 %
5 Tax data, previous year donor	1.8 %	0.3 %	1.0 %	5.1 %
6 Tax data, nearest neighbour donor	0.8 %	0.1 %	0.4 %	1.8 %
7 Unit non-response	8.0 %	1.4 %	5.4 %	26.3 %
All	100.0 %	100.0 %	100.0 %	100.0 %

The number of enterprises that are collected via direct inquiry or tax data without errors (internally consistent) is 84% of all enterprises. Direct inquiry is not considered as edit fail, because these are corrected manually and not flagged. According to the turnover the share of these enterprises rises to 95%. That is the turnover share of enterprises that fail in edits is only 5%. These enterprises can be divided into different edit-groups. The best quality groups are outlier detection, small errors and previous year donor (cold deck imputation). The cumulative turnover of these units is around 72% of the erroneous observations. Rest of the observations 1,5% of total turnover are edited using nearest neighbour imputation. These observations include also unit non-response.

The relative efficiency of the editing methods described above is a tricky issue. For 1998 data Statistics Finland launched a simulation to resolve the issue. This study was used to further develop the current practices and to decide the preference order for the editing methods.

## *5 Statistics Finland's direct inquiry and Tax authority's questionnaires*

The questionnaires used for Finnish SBS enterprise data are Statistics Finland's TILKES and Tax authority's 60, 61A, 61B and 62. The TILKES questionnaire includes only additional information for regular profit and loss account and balance sheet reporting. This additional data is used for producing data according to the EU Council Regulation 58/97 and National, Financial and Regional Accounts require-

ments. See the TILKES questionnaire  
[http://www.tilastokeskus.fi/keruu/yrti/tilkes05\\_en.pdf](http://www.tilastokeskus.fi/keruu/yrti/tilkes05_en.pdf).

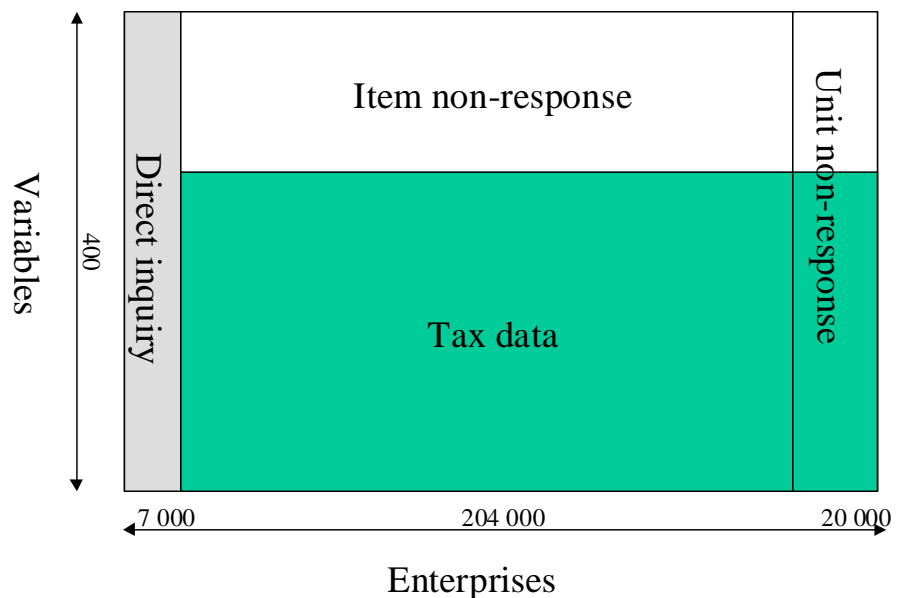
The Tax authority's form 60 includes the profit and loss account data. The forms 61A and 61B are balance sheets assets and liabilities. The questionnaire 62 is on investments and depreciation. For the year 2006 the questionnaire 6B has replaced the questionnaires 60, 61A and 61B, that is the old questionnaires are not available at Internet. See the Tax authority's questionnaires  
[http://www.vero.fi/default.asp?language=ENG&domain=VERO\\_ENGLISH](http://www.vero.fi/default.asp?language=ENG&domain=VERO_ENGLISH).

The establishment questionnaire (T5) is closely related to the enterprise level questionnaire TILKES. See the chapter 7 on establishment statistics.

## 6 Imputation methods

There are two types of imputation methods used in the Finnish SBS data. First type is donor imputation which is applied for unit non-response. Unit non-response arises if the data is not received either from direct inquiry or from the Tax authority data. For erroneous observations the donor imputation is sometimes used as a part of the editing process.

After direct inquiry and tax data are received and edited the imputation starts. Imputation process can be divided into two separate steps:



1) First step is unit non-response. Only variables included in tax data are imputed using nearest neighbour imputation with distance measure. Distance measure is presented in the chapter 6.2. Unit non-response includes those units that have not ended their accounting period during the year, extended accounting period, or that have not sent their accounting data to Tax Authority. Around 20 000 units are included in this group.

2) Second step is item non-response. Item non-response refers to mass imputation of the variables included in direct inquiry and not received from Tax Authority. Primary method used here is regression imputation with outlier detection and weighting if necessary.

## 6.1 Classifying observations

For the unit and item non-response imputations it is critical to define the correct and imputed observations (correct = C and imputed = I).

Categories are: UC, UI, IC and II. For unit non-response (U) the group of correct observations UC includes all received from the direct inquiry or that have passed the tax files' edits. On the other hand the group of imputed UI consist of the units which are edited using nearest neighbour method or are not received from direct inquiry or tax files being really unit non-response. Basically unit non response includes units in business register, but which are not received from tax files or direct inquiry.

For the item non-response the group of correct IC is defined as units received from direct inquiry. Thus the group of imputed (II) being defined as the units received from tax files or from the unit non-response imputation (UI).

## 6.2 Unit non-response

Basically the SBS survey in Finland is a census. Still there is a group of enterprises that do not give data to the tax authority or Statistics Finland. There are two reasons for this missing data. First some enterprises have not ended their accounting period during the statistical year, extended accounting period. Secondly, there is business registers over-coverage and enterprises that do not simply send the data to Tax Authority. The last group is rather numerous and consists of very small enterprises having only minor economic significance. That is, only first group is crucial.

Currently unit non-response is imputed using nearest neighbour imputation. This is a version of donor imputation, where donor is searched via minimisation function:

$$dist = \min \sum_{i \in V} |\ln(nr_i) - \ln(d_i)| \quad || \quad nr_i \in UI, \quad d_i \in UC$$

Dist = distance measure, nr = value for missing unit, d = value for the donor, i = variable for which the distance is calculated, V = vector of variables for which the distance is calculated (turnover and number of personnel), UI is group of unit non response and UC is group of correct observations.

The necessary data to compile the distance measure is downloaded from Business Register. The frame definition, turnover and personnel data is also available from Business Register. if income tax files are missing for a particular enterprise then Value-added tax files are used to estimate the missing data.

Unit non-response imputation is applied for profit and loss account and balance sheet variables that are available from tax data.

First the population of the possible donors is defined. Units that have first quality profit and loss account or balance sheet are feasible for donors. Imputations for these two sets of variables (profit and loss account and balance sheet) are separate. Then donors are searched first at national 5-digit level then group (3-digit) and finally 1-digit (not section) level. Donor is searched, if at least 50 units are found in the sub-population, if not then more aggregated level is chosen.

The minimum amount of units (=50) is crucial to have sufficient group of donors. Otherwise the method would be risky and limiting the set of possible donor values. It is paramount, that the donor group covers maximum number of feasible values before the selection.

### 6.3 Item non-response

Basically item non-response imputation is used to generate the contents of the direct inquiry for those enterprises that are not received via the direct survey. There are 120 questions in the form, which implies that are multiple item non-response regression models. Still all models can be presented in the generalised regression form.

$$y_i = a_i x_i + e_i \quad \| \quad i \in IC$$

The  $y_i$  is the dependent variable's value for the observation  $i$ . This is divided into two parts. The systematic part  $a_i x_i$  and the random part  $e_i$ . The systematic part can be further divided into  $a_i$  and  $x_i$  being called as the parameter and the explanatory variable. Here  $i$  is a unit from the group of correct observations (IC).

The used method to estimate the parameters of the equation is ordinary least squares (OLS) -regression. This implies that  $a_i$  can be estimated using the form.

$$a_{\text{hat}}_i = \sum x_i * y_i / \sum x_i^2 \quad \| \quad i \in IC$$

The method used in Statistics Finland is deterministic and does not include the random term  $e_i$ . The unavoidable consequence of this solution is, that the variation for the imputed variables is under estimated.

Population of the "correct" observations (IC) is not a random sample from the frame of the survey. Only larger enterprises are chosen to direct inquiry and this will bias the results of the imputations. To reduce the bias weighting is applied.

#### 6.2.1. Weighting the small units

Due to the fact the units received from tax files are rather small employing less than 20 persons weighting is crucial to reduce the bias otherwise generated by the imputation process.

The practical solution is to have larger weights for the smaller units from the direct inquiry (>20). The weight applied here is the inverse of the number of the employees (1/employees).

#### 6.2.2 Weighting the outliers

For the outlier weighting particular property of SAS software is used. The REG procedure includes DFFITS option which enables the user to define weights for those units that have excess effect for the parameter. In general feasible constraint is that the absolute value of DFFITS < 2. If DFFITS >= 2 then the observation is weighted by 0.1. Also all holding companies are excluded from the group IC.

#### 6.2.3 Activity classifications

The objective is to conduct imputations on the most exact activity level. In Finland this refers to the national activity classification (TOL2002) 5-digit level. The classification is equal to NACE 1.1 down to class level, but the 5-digit is national adaptation. Basically imputations start from the most disaggregated level towards more aggregated levels. If 15 or more observations are found, then imputations are conducted, else the process moves to more aggregated level. In principle the imputations starts from the 5-digit to the section level.

### 6.2.4 Details on turnover and other operating income

Turnover and operating income data are disaggregated for various components. These details are presented in the box below.

<i>Variable</i>	<i>Parameter 1</i>	<i>Parameter 2</i>	<i>Estimate</i>
Deliveries of products	T1	T11	P1
Deliveries of produced electricity	T1	T12	P2
Deliveries of produced heat	T1	T13	P3
Network activity	T1	T14	P4
Industrial repair and installation deliveries	T1	T15	P5
Paid labour	T1	T16	P6
Construction of buildings	T2	T21	P7
Land and water construction	T2	T22	P8
Renting the construction equipment with operator	T2	T23	P9
Other construction (heating and ventilation installations, painting etc.)	T2	T24	P10
Share trading in construction	T2	T25	P11
Wholesale and retail trade	T3	T31	P12
Commission trade	T3	T32	P13
Other unspecified turnover	T3	T33	P14

Two level regressions are used to compile the *Estimate*. In the first regression *Parameter 1* (T1 - T3) are estimated. The second regression further estimates *Parameter 2*. Finally the *Estimates* are estimated by multiplying *Parameter 1* and *Parameter 2* by rows respectively  $T1 * T11 = P1$ ,  $T1 * T12 = P2$  etc..

The itemisation of the operating income is more simple than for turnover. Operating income is divided to six components.

<i>Variable</i>	<i>Parameter</i>
Rental returns from land and water areas	T1
Other rental returns	T2
Royalties: returns from patents and licences	T3
Proceeds from sales of fixed assets	T4
Grants and subsidies from the State or local authorities	T5
Other operating income	T6

### 6.2.5 Details on expenses

The expenses disaggregation follows the logic of the turnover disaggregation. The expenses components in profit and loss account is shown below. Also disaggregation parameters are shown.

In Finnish profit and loss accounting scheme expenses are divided to 3 different categories. These are *Purchases of materials, implements and goods*, *External services* and *Other expenses*. The categories are further diagggregated using regression parameters *T1 - T20*.

<i>Variable</i>	<i>Cost Item</i>	<i>Parameter</i>
Acquisitions of materials and supplies (excl. Energy, packaging materials)	Purchases of materials, implements and goods	T1
Acquisition of merchandise (incl. purchases of electricity and heat acquired for resale)	Purchases of materials, implements and goods	T2
Acquisiton of packaging materials	Purchases of materials, implements and goods	T3
Acquisiton of fuels	Purchases of materials, implements and goods	T4
Acquisition of electricity for own use (excl. exchange electricity)	Purchases of materials, implements and goods	T5
Acquisition of heat for own use (excl. exchange heat)	Purchases of materials, implements and goods	T6
Contracted repair, maintenance and installation work	External services	T7
Contracted paid labour (incl. printing etc. in publishing)	External services	T8
Subcontracting	External services	T9
Renting of labour force	External services	T10
Research and development expenses	Other expenses	T11
Transport and storage expenses	Other expenses	T12
Advertising, sale and marketing expenses	Other expenses	T13
Computer, design and programming expenses	Other expenses	T14
Royalties: expenses from patents and licences	Other expenses	T15
Rents from land and water	Other expenses	T16
Leasing rents	Other expenses	T17
Other rents	Other expenses	T18
Losses from sales of fixed assets	Other expenses	T19
Other expenses not mentioned above	Other expenses	T20

### 6.2.6 Re-scaling the regression parameters

As described above the regression parameters are used for itemisation of particular aggregates. That is, the sum of the parameter groups for turnover, other operating income and expenses should add up to one. This is not usually the case and parameters need to be re-scaled to add up to one. This is done by dividing the parameters by the sum of parameters inside the group.

## 7 Establishment statistics (local KAU)

Establishment level statistics are compiled for the NACE 1.1 activities C - E. The data collection relies partly on the direct data collection and partly on the use of enterprise level data. The direct inquiry is collected from the establishments that belong to the enterprise employing more than 20 persons. The cut off sample for 2005 was 1841.

The establishment questionnaire is closely related to the enterprise level questionnaire TILKES. See the chapter 5 on questionnaires.

The enterprises below the threshold are assumed to be uni-regional and uni-industrial. That is, the enterprise level figures are accumulated to one reference establishment. The establishment is chosen using the proximity algorithm which takes into account the location, activity and size of the establishment.

## 8 Multi yearly trade statistics

Multi yearly trade statistics are compiled as separate surveys according to the requirements of the SBS regulation. All multi yearly trade statistics have been compiled once each. Therefore we have no fixed practises for the production, especially for the sample design and estimation.

### Survey

The survey frame has been the same as for the SBS data in retail and whole trade. For regional data the frame was based on the establishments of the business register.

All the data for retail trade and whole trade has been collected directly from the enterprises. For multi yearly regional data some data on sales space was collected by a private enterprise.

### Sampling methods

Sampling method has been stratified random sampling. Stratification is based on activity and size for retail and whole trade. In regional data area (NUTS2) is also used in stratification. All the biggest enterprises are surveyed.

Series	Ref. year	Sample size	Response rate
Retail trade	2002	2.289	69 %
Whole trade	2003	1.195	76 %
Regional data	2004	2.105	89 %
Car trade	2005	203	100%

### Data collection

The data has been collected using paper forms. Electronic data collection will be used in the car trade data collection in this autumn. Usually we send two reminders by post. The third reminder is by phone to the biggest enterprises. Penalties are not used.

## 9 Preliminary data compilation

The main methodological difference in the preliminary data compilation compared with the definite data is the unit non-response imputation. The preliminary data is compiled when approximately 65-75 percent of the units and over 90 percent of the total turnover is received. This implies that unit non-response imputation has to be applied to overcome the problems of the bias in the data.

The non-response imputation process is based on the use of the previous year data for the particular units. Also the data from VAT files are used to predict the change in the turnover. Some cost structure variables are adapted using regression coefficients, which are estimated using the received clean data.

### 9.1 Structural imputation coefficients

Before applying the turnover change coefficients the structural imputation coefficients are used for the structural change estimation. The idea is to modify the previous year for the non-response units using the information available from the received clean data.

The most important variables are modified using regression imputation. The formulas are presented below. The  $t$  indicates the Statistical year and  $t-1$  is the previous year.

- Turnover ( $t$ )=turnover ( $t-1$ )
- Purchases of the goods and services ( $t$ )= Purchases of the goods and services ( $t-1$ )
- Personnel costs ( $t$ )= Personnel costs ( $t-1$ )
- Depreciation ( $t$ )= Depreciation ( $t-1$ )
- Other expenses ( $t$ )= Other expenses ( $t-1$ )
- Financial income ( $t$ )=financial income ( $t-1$ )
- Financial expenses ( $t$ )= Financial expenses ( $t-1$ )

The same hierarchical activity algorithm as for the item non-response imputation is used, but the minimum number of the units for compiling the imputation starts now with 200 and ends to 100. That is, the tax files are used as clean data, so very extensive number of the clean observation are available.

### 9.2 Use of the VAT files

The VAT files are received on monthly basis, which can be used for the preliminary estimate for the turnover change. Using turnover change the business volume is approximated and the enterprise profit and loss account is scaled to the new estimated level.

## 10 Output and deliveries

Output and deliveries can be divided into two major components. First is to full fill the domestic requirements and second is to full fill Eurostat's (EU Commission) requirements. The first part is called Financial Statement Statistics (FSS) and the second is Structural Business Statistics (SBS). Both products are produced from the same database with same methodology, only the contents (variables and classifications) of the data differs.

## 10.1 Financial Statement Statistics (FSS)

These Statistics are divided into five products according to the industry breakdown and statistical unit. The enterprise and establishment level statistics are produced for industry (NACE 1.1 C-E). Only enterprise level statistics are produced for construction (NACE 1.1 F), trade (NACE 1.1 G), hotel and restaurant activities (NACE 1.1 H), transport (NACE 1.1 I) and business services (NACE 1.1 72, partly 74, 804, 851, 900 and 93 (laundry services)).

The FSS data is published at Internet using PX-web service and electronic reports at HTML format. For 2005 the preliminary data is published at T+6 months and definite at T+12 months.

## 10.2 Regional Statistics on Manufacturing

Regional Statistics on Manufacturing are published for the activities C - E. The data is published on internet. For 2005 the preliminary data is published at T+6 months and definite at T+12 months. The data is available also in English; [http://tilastokeskus.fi/til/atoi/index\\_en.html](http://tilastokeskus.fi/til/atoi/index_en.html)

## 10.3 Structural Business Statistics (SBS)

SBS data covers annexes 1-4 of EU CR 58/97 business services. The preliminary and definite series are delivered to Eurostat. Preliminary estimates are delivered at T+10 months and definite at T+18 months.

# 11 Benefits and challenges in using administrative data for business statistics

There are several pros and cons in using administrative data in the business surveys. Also the challenges needs to be discussed in more detail.

The main benefits are:

1. Reduced response burden
2. Reduced costs
3. Improved coverage
4. Improved precision (incl. more detailed presentation)
5. Enterprises have strong intensive to provide correct data (this relates to the threat of punishments)

And the challenges are:

1. Administrative data might be erroneous (data entry errors and inconsistent concepts)
2. Scope of information is limited
3. Statistical unit (enterprise) versus legal unit
4. Dependency on administrative data

Usually administrative data is very numerous including data entry errors. Also the definitions of the variables are not what statisticians would like to have. That is, some adjustments needs to be tailored to overcome these challenges. At Statistics Finland for SBS administrative data treatment the automated editing procedures are developed to correct numerous errors included in the administrative data. To transform the concepts and variables used in administrative data towards statistical concepts one needs the direct inquiry. The direct inquiry is used for adjusting the administrative data variables towards the statistical variables. Also all important statistical variables are not available from the administrative sources and the direct survey is needed to fill the gaps.

In general the unit used in administrative data is legal unit and the unit intended to use in statistics is the statistical unit (enterprise). To link the statistical units and legal units is one of the future challenges for the SBS at Statistics Finland.

Finally the statistics are depended on the administrative data. In the case the institutional concepts, questionnaires or time tables are adapted by administrative authorities also statistical institutions needs to adapt these changes. This implies close cooperation between administrative authorities. Being aware of all possible changes in legislation and questionnaires. Furthermore the data treatment tools for treating administrative data should be as generalised as possible. This helps to update the automated tools when the changes in the legislation or questionnaires will occur.