



STATISTIKOS DEPARTAMENTAS  
STATISTICS LITHUANIA

## SBS DATA PRODUCTION: ESTIMATION OF NON-EXISTENT VARIABLES



Antanina Valiulienė

Enterprise Statistics Division

25-29 June 2012

SBS data production. Estimation of non-existent variables



[www.stat.gov.lt](http://www.stat.gov.lt)





## DATA INTEGRATION

### ◆ DATA FROM ADMINISTRATIVE SOURCES AND ADDITIONAL STATISTICAL DATA SOURCES ARE USED FOR THE PRODUCTION OF FINAL SBS DATA:

- Update of the frozen list of active enterprises
- Estimation of small businesses
- Estimation of non-responding enterprises





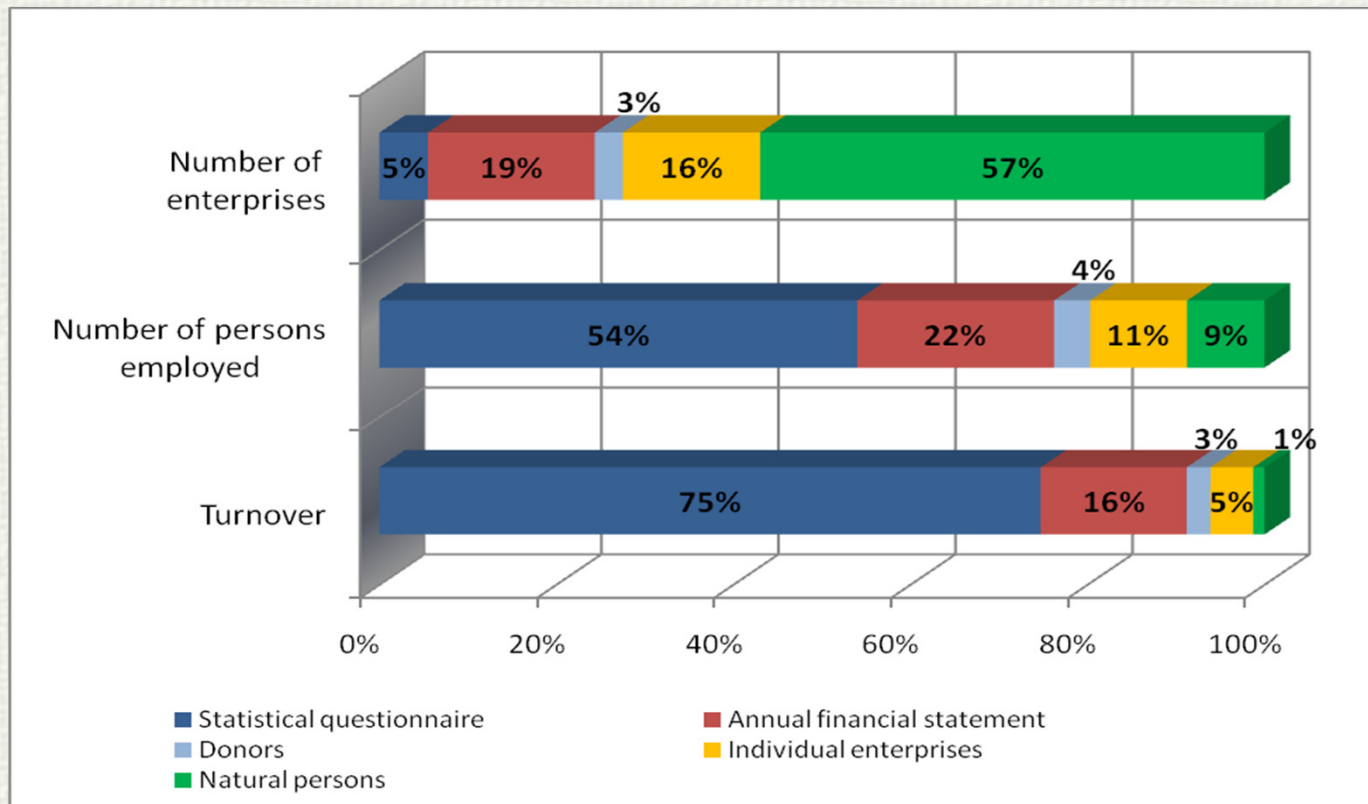
## DATA INTEGRATION (cont.)

### ◆ Data combination:

- **Admin Data** which directly correspond to the variables in the statistical questionnaire are directly transferred to the database (the number of employees, assets, equity, turnover, etc.).
- **Admin Data** which do not completely correspond to the variables in the statistical questionnaire, for example wages and salaries, are estimated and imputed.
- **Non-existent variables** are estimated and imputed.



## DISTRIBUTION OF THE MAIN SBS INDICATORS BY SOURCE







## ADMINISTRATIVE DATA PROCESSING

◆ The estimation are performed for following variables (group of variables)

- Employees
- Assets and equity
- Stocks
- Sales, costs, profit (loss)
- Gross investment in tangible goods
- Purchases of goods and services





## ESTIMATION APPROACHES

- ◆ **For the estimation and imputation of missing variables, the following estimation approaches are used:**
  - **Structural coefficients of the data of the same enterprise from the previous year**
  - **Donor values (nearest neighbour method) or the structural coefficients of the donor data**
  - **Structural coefficients calculated from the data of respondent enterprises (statistical questionnaire) grouped by various classes**
  - **Part of variables are calculated from the existing initial indicators (purchase of goods and services, investment)**





## ESTIMATION APPROACHES

- ◆ **As for enterprises which did not provide an annual financial statement but their turnover is known from other sources, the variables needed for the statistical questionnaire are estimated as follows:**
  - **The previous year's data of the same enterprise are multiplied by the alteration coefficient of turnover (used for small individual enterprises)**
  - **All data of the donor are imputed directly (the nearest neighbour selected by turnover, number of employees, total balance sheet and NACE code)**





## ESTIMATION APPROACHES USED FOR EMPLOYEES' VARIABLES

- ◆ All units (which are not surveyed by statistical questionnaire) are supplemented by available data on employees from administrative and statistical data sources (SODRA, STS and Quarterly Statistical Survey on Earnings).
- ◆ The variable *wages and salaries 13320* from SODRA is not the same as required by SBS since it does not include business trip daily allowances, cost-of-living, family allowances, etc. For this reason, the value of this variable is evaluated (increased) randomly.
- ◆ Non-existent variables are imputed by donor values.
- ◆ The number of employees in full-time equivalent units (16140) is calculated from primary (initial) variables.



## ESTIMATION APPROACHES USED FOR EMPLOYEES' VARIABLES (2)

Table. *Evaluation of SBS variables related to employees*

No. in SQ F01	SBS code	Name	Data sources or estimation methods used
100	16 11 0 16 13 0	Number of persons employed, Number of employees	Primary data from the State Social Insurance Fund Board (SODRA)
101	-	Number of part-time employees	Primary data for a subset of the population from the Quarterly Statistical Survey on Earnings; for the rest – donor values*
103	16 15 0	Number of hours worked by employees	Primary data for a subset of the population from the Quarterly Statistical Survey on Earnings and STS surveys; for the rest – donor values
104	13 32 0	Wages and salaries	Primary data from the State Social Insurance Fund Board (SODRA) randomly increased by coefficients
105	13 33 0	Social security costs	=104*0.31
106	-	Number of hours worked by part-time employees	Primary data for a subset of the population from the Quarterly Statistical Survey on Earnings; for the rest – donor values
-	16 14 0	Number of employees in full-time equivalent units	Data for a subset of the population from the Quarterly Statistical Survey on Earnings; for the rest – calculated from initial variables





## DONOR METHOD

- ◆ Donors are used to evaluate variables *number of part-time employees, number of hours worked by part-time employees and number of hours worked by employees (16150)*.
- ◆ The variables *number of persons employed (16 11 0)* and *wages and salaries (13 32 0)* from SQ F01 and SODRA are used to find the most suitable donors.
- ◆ Donors have to be in line with certain rules, and in each step the potential donor is selected.
- ◆ After the required variables is selected, the consistency between those variables is checked additionally. The inconsistency errors is edited automatically.





## ESTIMATION APPROACHES USED FOR ASSETS AND EQUITY

To estimate the missing/non-existent variables for assets and equity sections are used:

- ◆ **Structural coefficients of the data of the same enterprise from the previous year.**
- ◆ **If the previous year's data are not available, the structural coefficients of the donor's data are used. Donors are selected by the variable *number of persons employed* and by NACE code.**





## ESTIMATION APPROACHES USED FOR STOCKS

◆ As regards stocks, the following breakdown can be made:

- *stocks of finished goods*
- *stocks of work in progress*
- *stocks of goods and services purchased for resale in the same condition as received*
- *stocks of raw materials and consumables*

To estimate the missing/non-existent variables:

◆ Structural coefficients of the data of the same enterprise from the previous year are used.

◆ If the previous year's data are not available, then the missing variables are estimated by NACE code:

- For NACE codes 2–43, 49–96 (except for 6810), 452, 454, the *stocks* value is equal to *stocks of raw materials and consumables*
- For NACE codes 45–47 (except for 452, 454), 6810, the *stocks* value is equal to *stocks of goods and services purchased for resale in the same condition as received*





## **ESTIMATION APPROACHES USED FOR SALES, COSTS, PROFIT (LOSS)**

**To estimate the missing/non-existent variables:**

- ◆ **Logical edits are used during the editing procedure.**
- ◆ **Only a few variables are estimated.**





## ESTIMATION APPROACHES USED FOR GROSS INVESTMENT IN TANGIBLE GOODS

◆ For the estimation of *gross investment in tangible goods*, the following variables are used:

- *Tangible assets at the beginning of the year*
- *Tangible assets at the end of the year*
- *Depreciation*

◆ The variable *gross investment in tangible goods* is calculated as follows:

$$\begin{aligned} X = & + \textit{Tangible assets at the end of the year} \\ & + \textit{Depreciation} \\ & - \textit{Tangible assets at the beginning of the year} \end{aligned}$$

If  $X > 0$ , then  $X = \textit{gross investment in tangible goods}$

If  $X < 0$ , then  $X = \textit{disposal of tangible goods}$





## ESTIMATION APPROACHES USED FOR PURCHASES OF GOODS AND SERVICES

- ◆ *The variable purchases of goods and services is calculated from the existing initial variables:*
  - + a sum of all costs and expenses
  - depreciation and amortisation, employee benefits and operating taxes
  - + changes in stocks
- ◆ **For the rest of variables (*purchases of: goods and services for resale, raw materials and consumables, fuels, electricity, heat, etc.*), structural coefficients of the same year's data are used.**
  - **The coefficients are calculated according to NACE 2-digit level code and the number of persons employed. According to the number of persons employed, the enterprises are divided into 2 groups: 1–49, 50 +.**