Frame Error Impact on Structural Business Statistics Surveys

FIFTH INTERNATIONAL WORKSHOP ON
BUSINESS DATA COLLECTION METHODOLOGY

19 - 21 September 2018 - Statistics Portugal, Lisbon

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Statistics
Netherlands

Statistisk sentralbyrå
Statistics Norway

University of Ljubljana
FACULTY OF ECONOMICS

United States
Census
Bureau

Frame Error Impact on SBS Surveys in KAS

19 - 21 September 2018 - Statistics Portugal, Lisbon

Content

- Introduction
- Methodology
  – sampling
  – collection of data
- Quality
- Frame error
- Analysis
- Conclusions
- Questions

Introduction

- Business surveys are often the only data source used to collect economic information about enterprises.

- In particular, Structural Business Survey (SBS) are based on sample data and yearly produce aggregated economic data such as turnover, production value, personnel costs, etc.

Methodology

The sample frame consists of a list of active enterprises maintained in the Statistical

Business Register (SBR) within KAS. According to the methodology used in statistic, active enterprises are considered only those who have declared one of the taxes or have declared any employees during the reference year.

- This paper aims to identify and measure the impact of principal frame errors on the sample-based estimations using auxiliary administrative variables

Quality

- The SBS is based on the use of the Statistical Business Register (SBR) as the reference frame.
- Quality of statistics produced by each survey is related to the quality of the SBR.

Auxiliary administrative information for turnover: the fiscal turnover

- The use of administrative sources for statistical purposes continues to be one of the strategic purposes of any statistical institutes.

- But the possibility to substitute direct information with available administrative data is dependent, where the needed information exist, on their quality (data from Tax authority of Kosovo).

Types of errors

● Several types of error affect estimates;

 - it is possible to distinguish between sampling and non-sampling errors.

Frame errors and their impact

- It is known that the purpose of each survey is to produce estimate as accurate as possible of a given unknown parameter.

- In the estimation process of economic aggregates of the SBS survey, both frame and non- response errors are treated as non-responses and measures of the overall error are supplied in the production of statistics.

Cont.

- The impact of non-sampling errors is more difficult to measure. Among non-sampling errors frame errors and non-response errors, represent the major sources of errors.

Cont.

- Sampling errors can be evaluated according to some measures supplied by the sampling theory for probability sampling

Cont.- 2

- The level of errors in the register and the errors in the sampled-based estimation are correlated.

- The accuracy of estimates depends on their variability and bias.

- Their magnitude determines the overall error.

Frame and sample size

- In SBS survey for 2016, the frame of the survey was 36880 units, from which have been taken for sample 3151, which represent 8.54 % of all active enterprises.

Some findings regarding the SBS survey in KAS

- Sampling and survey coverage

  - the general rule is to cover at least 80 percent of activity, notably 80 percent of turnover from business register.

Levels of stratification

- (i) first stratification level - by activity NACE Rev 2. by section,

- (ii) second stratification level - by size (initially three strata of size class by activity),

- (iii) third stratification level was within 4 digit by size within the third class.

Analysis – some figures

- In the process to estimate data from the survey, an important step is the analysis of information from economical point of view.

- During the analysis we found some illogical data as following:

- Expenditure were higher than turnover;

- Wages and salaries for instance in some cases 50 euro/employee

- Turnover per employee 1000 euro whereas wages per employee more than 1200 euros, in such way those figures didn't make any sense.

- In such cases we have used comparative method within the same sector for different enterprises and Comparative method in different time for the same enterprises also.

- Intersection analysis

Example: 500 = purchasing; 400 = Turnover; 300 = salaries; 200 = number of employees

Stat. Units	NACE code	Turnover	Purchasing	Salary	No. of employee	Ratio 4/5	Av. Salary
A	46	3,452,165	2,456,152	75,850	25	0.71	252.83
B	46	16,356,145	10,556,085	95,851	30	0.65	266.25
C	46	15,467,154	10,587,095	85,851	25	0.68	286.17
D	46	12,158,250	17,850,950	72,850	18	1.47	337.27
E	46	8,956,985	6,857,599	10,500	15	0.77	58.33
F	46	589,950	256,355	158,500	27	0.43	489.20
total		56,980,649	48,564,236	499,401	140	0.85	297.26
*taken into account		44,822,399	30,713,286	340,901	98	0.69	289.88

Looking at this we have three cases with illogical data:

Enterprise "D" the cost of buying goods it's higher than Value of Sale or 17,850,950.0 > 12,158,250. In this case we have observed that cost of buying goods has been exaggerated. What we did? We used Average method within the section.

$$\overline{X} = \frac{\sum Xi}{\sum Yi} = \frac{448223991}{307132860} = 1.45$$

Xi = turnover

Yi = purchasing

Based on the result of formula the ratio of Purchasing to Turnover should be 0.69 and not 1.45, after that we corrected the value from 17,850,950.0 to 8,389,192.5 to prove 8,389,192.5/ 12,158,250 = 0.69*100 = 69%

Notice: were from we got the number 8,389,192.5   0.69* 12,158,250 = 8,389,192.5 euro

After the adjusted data the table will look like this:

Stat. Units	NACE code	Turnover	Purchasing	Salary	No. of employee	Ratio 4/5	Av. Salary
A	46	3,452,165	2,456,152	75,850	25	0.71	252.83
B	46	16,356,145	10,556,085	95,851	30	0.65	266.25
C	46	15,467,154	10,587,095	85,851	25	0.68	286.17
D	46	12,158,250	8,389,192.5	72,850	18	0.69	337.27
E	46	8,956,985	6,857,599	10,500	15	0.77	58.33
F	46	589,950	256,355	158,500	27	0.43	489.20
total		56,980,649	39,102,479	499,401	140	0.85	297.26
*taken into account		44,822,399	39,102,479	340,901	98	0.69	289.88

Main outcomes

- number of employees;

- turnover;

- value of purchases and detail of these purchases;

- value of the inventories at the beginning of the year and at the end of the year;

- value of the taxes paid by enterprises;

- value and details concerning the investment;

Actions to speed up or increase the response rate have been adopted:

- enterprises on delay are subjected to one reminder by post and one by phone.

- This survey (as the other SBS and most of the STS surveys) is managed in a centralised statistical information system aiming to assure also a rotating system among sampled units to reduce burden on enterprises.

Conclusion

- The unsatisfactory sampling survey response rate together with the availability of a huge amount of data from administrative sources (balance sheets and tax data) has suggested some adjustments in the SBS production process.

Cont.

- The integration of the original SBS sample with administrative sources has allowed both to increase the response rate and to measure the discrepancies in the final estimation due to unit non-response.

- Using the comparative method had shown a good result for improvement of quality of data.

- The estimation (grossing up) the data of SBS result to level of the economy we used method of coefficient which have shown good result, for all outcomes (indicators),

- A further work needs to be done like a more disaggregated analysis (NACE at 4 digits, NACE at 3 digits and size classes)
- A further analysis on the informative contents of tax data could permit to extend this experiment to other SBS variables.

"To err is human, to forgive divine – but
to include errors in your design is statistical.

Kish, 1977

Thank You for your attention!

Any Questions?

hysni.elshani@rks-gov.net

hysni.el@gmail.com