Study on different sampling methods for Structural Business Survey at the municipal level

Liljana Boçi
Institute of Statistics Albania

Abstract

This is a study on SBS-Survey, one of the main surveys of the national statistical program of Albania. The main objective of it is to suggest improvements on the sample design of SBS with the aim to use it on level of municipalities. The hypotheses included in this study are based on empirical analysis, where the information used is limited, compared to the traditional method used in the business structural survey. Suggestions for alternative, indirect methods are well provided and are suggested to be used on the future.

Keywords: Target Group, Stratification, Frame, Neyman allocation, Exhaustive, Accuracy level, Publication domain, Systematic selection, Reweighting, Implicit, Explicit.
.

Introduction

Different methods can be used for SBS sampling techniques despite the difficulties we may encounter due to the lack of auxiliary information, in order to ensure a good sample allocation and to guarantee accuracy in municipal level according to the new administrative division. Although being practical and objective, indirect methods are not always very accurate. Among them, probably the combination of the current method, by adding in stratification the municipal level, and, on the other hand, considering the number of employees taken from business register (instead of turnover estimates from previous year) can give us an identification of the economic situation at the municipal level, which nevertheless remains to be seen.

Whichever method chosen to study structural business data at municipal level, it will always reveal different levels of accuracy and precision, as well as different cost–benefit ratios and advantages and disadvantaged.

A good choice of a survey method requires clear research objectives and the awareness about the method efficiency and limitations towards the desired objectives. Limitations may vary as much as from operational conditions (lack of human recourses, longer time for field work management, time added for data cleaning), such as accessibility to the area or the interaction with local people (in the respective municipalities), to the project execution timing, or the available budget.

In this sense, this study aimed at investigating the strengths and weaknesses of these three methods; (1) Detailing further the actual stratum by adding the municipal level (with a grouping of micro enterprises 1-4 employees) and the required accuracy at the publication domain (according to the method applied so far in SBS), (2) detailing the stratum as in case one, but the required accuracy for this case is at the municipal level, (3) detailing the stratum as in case (1) and (2), but the required accuracy for this case is at the municipal level and publication domains according to current SBS.

ISSN 2519-1284
Acces online at www.iipccl.org

European Journal of Economics, Law and Social Sciences
IIPCCL Publishing, Graz-Austria

Vol. 5 No. 1
January, 2021

The three methods above, together with a cost-benefit analysis, give us a clearer picture of their effectiveness during practical implementation. Our specific objectives during this study were: (1) to compare the methods between them and to see which is easier to apply, this is due to the technical complication of sampling methodology, which directly affects statistical quality, (2) Verify which of the methods has the highest degree of accuracy, (3) Assess and compare the cost-benefit of each method.

Material and methods

The experimental analysis was carried out based on the register of enterprises that were active until 31 December 2016, BR2016. As a classification system for the economic activity of enterprises, it is used Nace rev 2 classification. Initially, a target group was defined, which includes all the enterprises with activity within structural business survey coverage. In all three cases, for frame construction it is used stratification method by; municipal level, publication domains as used so far in SBS, the size of the enterprise based on the number of employees, and the sample/exhaustive group. In this study, enterprises with 1 to 4 employees are grouped together, unlike the current SBS approach where enterprises with 1 employee were treated as a separate group. Enterprises with 10+ employees are surveyed entirely (exhaustively), just as in the current method. As a method of allocation, the "Neyman allocation" method was used based on the variable of the number of employees taken from business register (in the absence of the turnover value both in the register and as an estimated value from SBS of the previous year, the latter is due to more detailed strata at the municipal level). The accuracy in this case is required at the publication domain level in the whole country level, according to the current method implemented so far in SBS, with a precision starting from 1.6%.
In the second case, stratification remains the same as in the first case. The same allocation method is also used, with the only difference here that precision accuracy is required at the municipal level in the whole country level, starting again with 1.6%. In the third case, stratification remains the same as in the two above cases. The allocation method is the same as well. The only difference that has been applied in this method compared to the first two methods lies in the accuracy precision required at the municipal and publication domain level in the whole country level, starting again with 1.6%.

Results

The results we obtained from each method, for the sample size according to the respective level of accuracy and required quality assurance are as follows: (1) in the first method we required accuracy at the publication domain in the whole country level as it has been implemented so far in SBS, with a precision starting from 1.6 and the total sample size reaches to 33,494 enterprises (Table 1); (2) in the second method we required accuracy at the municipal level with a precision starting again from 1.6, and the sample size according to this method reaches to 33,351 enterprises (Table 2); (3) During the third method, we required accuracy at the municipal level and

ISSN 2519-1284
Acces online at www.iipccl.org

European Journal of Economics, Law and Social Sciences
IIPCCL Publishing, Graz-Austria

Vol. 5 No. 1
January, 2021

publication domain in the whole country level, with a precision starting again from 1.6, and the sample size reaches 61,970 enterprises (Table 3)

Discussion

A sample increase it is suggested by each of the method compared to the current sampling method, due to a fine detailed stratification, but also because of the level at which statistical accuracy is required. Some advantages and disadvantaged could be perceived in three methods, prior of making the final decision.

The first method resulted with a sample increase up to 33,494 enterprises as shown in Table 1. Advantages: The stratification is consistent with the current SBS method, adding only the fine detailed stratification by the new municipalities; the accuracy is required in the publication domains in the whole country level as well as it has been implemented so far in the current SBS method. Despite that number of employees (from BR2016) is used as the allocation variable, the method used remains the same, Neyman allocation method. The required accuracy in the publication domain level is in line with the purpose of the structural business survey and based on its regulation, recommended by Eurostat. Regardless of the level at which statistics accuracy is required, by having the detailed stratum at municipal level, we may implicitly guarantee a good representation in 61 newly created municipalities. This is realized during sample selection, initially sorted by stratum and then systematically select the sample according to the size specified in each of them. The large enterprises with 10+ employees are thoroughly surveyed in line with the current method.

Table 1. First method, precision required in publication domain level according to SBS

bg	stratum	storan	allvar	nhsh	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10
0709	0107091S	7	14	7	7	7	7	7	7	7	7	7	7	7
0709	0107093S	2	11	1.4142136	2	2	2	2	2	2	2	2	2	2
10A0	0110A01S	1	1	0	1	1	1	1	1	1	1	1	1	1
10A0	0110A04E	1	9	0	1	1	1	1	1	1	1	1	1	1
10B0	0110B01S	3	7	4.5825757	3	3	3	3	3	3	3	3	3	3
10B0	0110B03S	2	10	0	2	2	2	2	2	2	2	2	2	2
10B0	0110B04E	3	147	179.5745	3	3	3	3	3	3	3	3	3	3
.	.													
.	.													
6061	5060614E	50	6880	10574.541	50	50	50	50	50	50	50	50	50	50
6263	5062631S	691	925	522.65986	116	105	94	86	79	72	66	62	57	53
6263	5062633S	43	286	63.56062	20	20	20	20	20	20	20	20	20	20
6263	5062634E	45	1329	1344.5549	45	45	45	45	45	45	45	45	45	45
6800	5068001S	357	473	255.47372	70	64	58	53	49	45	41	38	36	33
6800	5068003S	35	223	43.355575	20	20	20	20	20	20	20	20	20	20
6800	5068004E	31	620	544.2244	31	31	31	31	31	31	31	31	31	31
6900	5069001S	1640	2392	1325.8235	234	210	190	173	158	144	132	122	113	104
.	.													
.	.													
6900	6169001S	2	2	0	2	2	2	2	2	2	2	2	2	2
7100	6171001S	1	1	0	1	1	1	1	1	1	1	1	1	1
7882	6178824E	1	153	0	1	1	1	1	1	1	1	1	1	1
8500	6185001S	1	1	0	1	1	1	1	1	1	1	1	1	1
8688	6186881S	1	1	0	1	1	1	1	1	1	1	1	1	1
9300	6193001S	2	2	0	2	2	2	2	2	2	2	2	2	2
9300	6193003S	1	5	0	1	1	1	1	1	1	1	1	1	1
9500	6195001S	3	3	0	3	3	3	3	3	3	3	3	3	3
9600	6196001S	10	10	0	10	10	10	10	10	10	10	10	10	10
		122,950	495,009		33,494	33,141	32,847	32,609	32,405	32,232	32,080	31,957	31,846	31,751
				precision	1.6	1.8	2	2.2	2.4	2.6	2.8	3	3.2	3.4

Disadvantages: A very large number of strata compared to the current method (4,379 strata), which makes their technical handling during the reweighting process difficult,

turning it into an almost impossible one. Such detailing risk the statistical quality of each stratum, as we reach below the limit of the elements we seek to be within. This is the criterion that applies during the current method. Also, this way would jeopardize their existence, as they would lose incase there would be no-response, out of scope, closed down or dormant enterprises between them. Despite the fact that the accuracy is required at the same level as in the current method, thus in publication domains, we are unable to guarantee statistical quality explicitly and according to the municipal level. Such a method raises the problem of time, the need for a longer period of time to gather information, it adds the need for people to regularly check the data, it takes longer to clean up the data, analyze the data, and process of the results as well. And the biggest risk to the implementation of this method has to do with: inclusion of a larger number of micro enterprisesin a sample, since the exhaustive part is defined by threshold and it is fixed. It requires a big budget compared to the current method.

Table 2. Second method, precision required in municipal level according to the new division

bg	stratum	storan	allvar	nhsh	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10
01	0107091S	7	14	7	7	7	7	7	7	7	7	7	7	7
01	0107093S	2	11	1.414213562	2	2	2	2	2	2	2	2	2	2
01	0110A01S	1	1	0	1	1	1	1	1	1	1	1	1	1
01	0110A04E	1	9	0	1	1	1	1	1	1	1	1	1	1
01	0110B01S	3	7	4.582575695	3	3	3	3	3	3	3	3	3	3
01	0110B03S	2	10	0	2	2	2	2	2	2	2	2	2	2
01	0110B04E	3	147	179.5744971	3	3	3	3	3	3	3	3	3	3
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
50	5060614E	50	6880	10574.54105	50	50	50	50	50	50	50	50	50	50
50	5062631S	691	925	522.6598611	20	20	20	20	20	20	20	20	20	20
50	5062633S	43	286	63.56061973	20	20	20	20	20	20	20	20	20	20
50	5062634E	45	1329	1344.554928	45	45	45	45	45	45	45	45	45	45
50	5068001S	357	473	255.4737173	20	20	20	20	20	20	20	20	20	20
50	5068003S	35	223	43.35557499	20	20	20	20	20	20	20	20	20	20
50	5068004E	31	620	544.2244022	31	31	31	31	31	31	31	31	31	31
50	5069001S	1640	2392	1325.823455	20	20	20	20	20	20	20	20	20	20
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
61	6169001S	2	2	0	2	2	2	2	2	2	2	2	2	2
61	6171001S	1	1	0	1	1	1	1	1	1	1	1	1	1
61	6178824E	1	153	0	1	1	1	1	1	1	1	1	1	1
61	6185001S	1	1	0	1	1	1	1	1	1	1	1	1	1
61	6186881S	1	1	0	1	1	1	1	1	1	1	1	1	1
61	6193001S	2	2	0	2	2	2	2	2	2	2	2	2	2
61	6193003S	1	5	0	1	1	1	1	1	1	1	1	1	1
61	6195001S	3	3	0	3	3	3	3	3	3	3	3	3	3
61	6196001S	10	10	0	10	10	10	10	10	10	10	10	10	10
		122,950	495,009		33,351	33,009	32,715	32,469	32,252	32,069	31,907	31,769	31,644	31,538
				precision	1.6	1.8	2	2.2	2.4	2.6	2.8	3	3.2	3.4

The second method resulted with a sample increase as well, up to 33,351 enterprises as shown in Table 2, but comparable to the size of the first method. Advantages: The level of accuracy is required at the municipal level. The same allocation method applies as in the first and the current one. Large enterprises with 10+ employees are thoroughly surveyed as well. Disadvantages: There are still a large number of strata (4,379 strata),most of them with very few elements, which would jeopardize the accuracy of the statistics we want to measure, the same situation as in the first method. On the other hand, although the required accuracy is at the municipal level, we are not able to guarantee statistical quality explicitly and according to the publication domains that are object of study for SBS. As a result, in this way we divert in meeting the recommendations fulfillment of Eurostat' SBS regulation. In addition, we lose

comparability with the current method. This method also requires a longer period of time for concluding all statistical processes, compared to the current method. There is also an increase in the number of micro enterprises in the sample, and consequently their involvement would jeopardize the statistical process and would result in data loss for some municipalities, the same situation as in the first method. This method requires a large budget as well, compared to the current method.

Table 3. Third method, precision in municipal and publication domains level

bg	stratum	storan	allvar	nhsh	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10
010709	0107091S	7	14	7	7	7	7	7	7	7	7	7	7	7
010709	0107093S	2	11	1.414213562	2	2	2	2	2	2	2	2	2	2
0110A0	0110A01S	1	1	0	1	1	1	1	1	1	1	1	1	1
0110A0	0110A04E	1	9	0	1	1	1	1	1	1	1	1	1	1
0110B0	0110B01S	3	7	4.582575695	3	3	3	3	3	3	3	3	3	3
0110B0	0110B03S	2	10	0	2	2	2	2	2	2	2	2	2	2
0110B0	0110B04E	3	147	179.5744971	3	3	3	3	3	3	3	3	3	3
.	.				.	.	.	.	.	.	.	.	.	.
.	.				.	.	.	.	.	.	.	.	.	.
506061	5060614E	50	6880	10574.54105	50	50	50	50	50	50	50	50	50	50
506263	5062631S	691	925	522.6598611	144	130	119	109	100	92	85	78	73	68
506263	5062633S	43	286	63.56061973	20	20	20	20	20	20	20	20	20	20
506263	5062634E	45	1329	1344.554928	45	45	45	45	45	45	45	45	45	45
506800	5068001S	357	473	255.4737173	112	103	96	89	82	76	71	66	62	58
506800	5068003S	35	223	43.35557499	20	20	20	20	20	20	20	20	20	20
506800	5068004E	31	620	544.2244022	31	31	31	31	31	31	31	31	31	31
506900	5069001S	1640	2392	1325.823455	294	266	242	220	202	185	171	157	146	135
.	.				.	.	.	.	.	.	.	.	.	.
.	.				.	.	.	.	.	.	.	.	.	.
616900	6169001S	2	2	0	2	2	2	2	2	2	2	2	2	2
617100	6171001S	1	1	0	1	1	1	1	1	1	1	1	1	1
617882	6178824E	1	153	0	1	1	1	1	1	1	1	1	1	1
618500	6185001S	1	1	0	1	1	1	1	1	1	1	1	1	1
618688	6186881S	1	1	0	1	1	1	1	1	1	1	1	1	1
619300	6193001S	2	2	0	2	2	2	2	2	2	2	2	2	2
619300	6193003S	1	5	0	1	1	1	1	1	1	1	1	1	1
619500	6195001S	3	3	0	3	3	3	3	3	3	3	3	3	3
619600	6196001S	10	10	0	10	10	10	10	10	10	10	10	10	10
		122,950	495,009		61,970	60,209	58,593	57,080	55,660	54,351	53,110	51,951	50,855	49,824
				precision	1.6	1.8	2	2.2	2.4	2.6	2.8	3	3.2	3.4

A considerable increase in the sample size was observed during the realization of the third method, compared with the first two methods mentioned above and the current SBS method. The total size of the sample goes up to 61,970 enterprises as shown in Table 3. Advantages: The stratification method performed is the same as we did in the first two methods. The allocation method is also the same. Accuracy required in this case was at the municipal level and the publication domain in the whole country level, which would ensure comparability with the current SBS method and would be in compliance with the recommendations of Eurostat'SBS regulation.Disadvantages: Again the number of strata remains high compared to the current method (4,379 strata). As we pointed out above in the first two methods, such detailing risk the statistical quality of each stratum (strata) and it is not what sampling theory suggests. The stratums we are talking about do not have enough observation elements, which will be asked later for the SBS purposes.It is worth re-emphasizing that this process technically would subsequently result in a very complicated re-weighting process, as the total number of strata exceeds the manageable limit of them (in number) for a survey. Actually, in the current method, the maximum number of strata during the re-weighting process is approximately 350. The threshold of exhaustive enterprises remains the same as in the previous methods mentioned above and in accordance with the current method. Here, we are dealing with an even greater increase in the number of micro enterprises included in the sample (almost twice as much as in the first two methods). Adding them would jeopardize the process of producing statistical data

ISSN 2519-1284
Acces online at www.iipccl.org

European Journal of Economics, Law and Social Sciences
IIPCCL Publishing, Graz-Austria

Vol. 5 No. 1
January, 2021

at the municipal level. Also, this method requires much longer time than the above mentioned methods to conclude the entire statistical process. This method requires much larger budget than the first two methods and the actual one, too.

Indirect methods were also proved but not presented in this material, because the obtained results do not meet the SBS survey objectives. Comparability with the current method is lost and it is not in line with the regulation recommended by Eurostat because in stratification process, the level of economic sections was used (instead of the publication domains) and accuracy was requested at the section level and asa combination of two (section with municipal level). While the structural business survey intended to guarantee quality and produce statistical data in the publication domains as this regulation suggests. Despite the strengths and weaknesses encountered in each of the methods discussed above and versus one another, we should note that: "All three methods jeopardize the quality of statistical dataproductionon the structural business survey in the requesteddetail for this survey, and in the fulfillment of the purpose thatsurvey has in itself according to Eurostat regulation!"

An effective way to collect structural business data at municipal level would be the usage of administrative resources. So, we would suggest to study as an alternative the fulfillment of SBS objectives through the administrative information that is declared by them. That's because such a largesamples that tend to go to census do not make sense to be suggested as solutions and they are not effective. Currently, some of the EU member states use administrative resources to publish structural business data.

Other alternative methods remain to be seen. Data estimation models in "Small area estimation" remain as an alternative to be explored. This is a method that requires a special study and deepened expertise regarding its technical applications, to see how effective it is as a method and at what cost it results.

References

Eurostat, SBS Regulation No.58/97, 20 December1996 amended, by European Regulation No. 295/2008 and two implemented acts (250/2009 and 251/2009).
INSTAT, Business Register of active enterprises for2016.
Sharon L. Lohr, Sampling design and Analyses,2ndedition.