Identification of Factors Affecting to Non-Sampling Errors: Study on Economic Census 2013/14

Srimali KGAA[1]
Wickramasinghe WMPM[2]
Aponsu GMLM[3]

ABSTRACT

Non-sampling error is a major problem arises in a survey and that directly affects to the survey results. They can be categorized as coverage (or frame) errors, non-response errors, measurement errors, data handling errors, data entering errors, data coding errors and data rewriting errors, and so on. This study attempts to examine the factors affecting to non-sampling errors based on Economic census 2013/14conducted by the Department of Census and Statistics (DCS). The dependent variable of the study was the changing of Sri Lanka Standard Industry Classification (SLSIC) code and the independent variables are the District, DS Division, Sector, Legal status, Nature, Registered institution, Location, Account maintenance, Main decision maker and No. of Employees. Correlation analysis and binary logistic regression analysis were used to analyze the data and it was found that the changing of SLSIC code was affected by sector, nature, location, gender, no. of employees and legal status.

KEYWORDS: Department of Census and Statistics (DCS), Economic Census, Factors, Non-Sampling Errors, Sri Lanka Standard Industry Classification Code (SLSIC)

INTRODUCTION

Economic census is a process of collecting, compiling, analyzing and disseminating the data related to economic activities carried out with the country's territory.

The first economic census was conducted by the industry trade and services division of the Department of Census and Statistics in 2013/1014. The subject areas considered in this survey are Milling and Quarrying, Manufacturing, Electricity, Gas Production and Supply, Water supply, Sewerage and Remediation Operations, and Construction Industry.

[1]Graduate, Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka

[2]Senior Lecturer, Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka

[3]Lecturer, Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka

DCS has developed the Sri Lanka Standard Industry Classification code (SLSIC) based on the International Standard Industrial Classification.

Non-sampling errors can exist when an obtained sample differs from the original selected sample. Non-sampling errors can be categorized as coverage (or frame) errors, non-response errors, measurement errors, data handling errors, data entering errors, data coding errors and data rewriting errors. Coverage errors occurred in incomplete frames of the survey. Measurement errors can be defined as the difference between the required information and the produced information. The non-response error is the most important type of error in the designing of the survey.

RESEARCH OBJECTIVE

The main objective of this research is to identify the factors affecting to non-sampling errors in the Economic census 2013/2014.

LITERATURE REVIEW

Assael & Keon (1982) described the minimization of the total error of non-

sampling and sampling errors and its components and the increment of the validity and consistency of the survey data. In this research, the components of total survey errors for several research designs and data collection methods are examined and found that Non-sampling errors are the major contributors to the total survey error. Jeremiah (2003) described three types of non-sampling errors; specification errors, non-response errors, and coverage or frame error.

Bond (1992) discovered non-sampling error issues in the survey of industrial research and development. This research discuss the behavior of the Industrial Research and Development, to control and measure non-sampling errors in a survey. In this study non-sampling errors are grouped in to five categories. Those categories are specification, coverage, response, non-response, and processing.

Roberts (2007) stated about mixing modes of data collection in a survey. It structured in the range of mode choices available to survey researchers and their advantages and disadvantages with respect to a range of criteria and including their impact on data quality. Using a combination of data collection modes, the survey designers are exploiting the potential offered. This research describes how people respond and focus in particular problem that leads to measurement error in a survey question.

METHODOLOGY

This study was carried out to find the factors affecting to non-sampling errors by considering 28385 data of economic census 2013/14. Descriptive analysis was used to compare variables and identify the behavior of the variables. Chi-square test was done to identify the relationship between dependent variable and independent variables separately. The Binary logistic regression method was used to identify the factors.

RESULTS AND DISCUSSION

The summary of the Changing of SLSIC Code-wise distribution is shown in Figure 1.

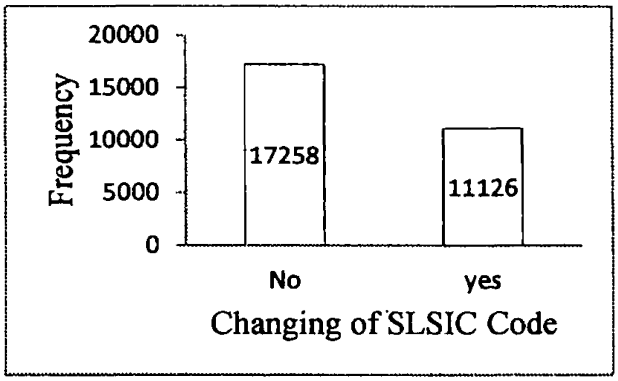

Figure 1: Changing of SLSIC Code

As indicated in Figure 1, it is clear that in most of the cases the SLSIC code has not been changed.

Summary of the changing of SLSIC code wise main decision maker distribution is given in Table 1.

Table 1: Changing of SLSIC Code wise Main Decision Maker

Main decision maker	SLSIC Code changed	
	No	Yes
Male	13159(58.9%)	9175(41.1%)
Female	4099(67.8%)	1951(32.2%)
Total	17258	11126

According to the Table 1, most of the males changed the SLSIC code in filling the questionnaire than females.

Summary of the changing of SLSIC code wise nature distribution is given in Table 2.

Table 2: Changing of SLSIC Code wise Nature

Nature	SLSIC Code changed	
	Yes	No
Single units	10020(37.2%)	16932(62.8%)
Head office	220(88%)	30(20%)
Multiple units	894(74.9%)	299(25.1%)
Total	11134	17261

As indicated in Table 2, higher percentage of head office changed SLSIC code than single units (without subdivisions) and multiple units (with subdivisions).

Chi-square test was used to identify the relationship between dependent variable and independent variable separately and it is represented in Table 3.

H_0: There is no relationship
H_1: There is a relationship

Table 3: Test for Correlation

Relationship between	P value	Result
District and Changing of SLSIC Code	0.000	Reject H_0
DS Division and Changing of SLSIC Code	0.000	Reject H_0
Sector and Changing of SLSIC Code	0.000	Reject H_0
Legal status and Changing of SLSIC Code	0.000	Reject H_0
Nature and Changing of SLSIC Code	0.000	Reject H_0
Registered institution and Changing of SLSIC Code	0.000	Reject H_0
Location and Changing of SLSIC Code	0.000	Reject H_0
Account and Changing of SLSIC Code	0.000	Reject H_0
Gender and Changing of SLSIC Code	0.000	Reject H_0
No. of employees and Changing of SLSIC Code	0.000	Reject H_0

According to Table 3, it is clear that all independent variables have a significance relationship with the changing of SLSIC code.

The binary logistic regression was used to identify the factors affecting to non-sampling error. The model summary is shown in Table 4.

Table 4: Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	37080.900[a]	.032	.044
2	36711.675[a]	.045	.061
3	36541.023[a]	.051	.069
4	36476.399[a]	.053	.072
5	36430.003[a]	.054	.074
6	36403.677[a]	**.055**	**.075**

As indicated in Table 4, the Cox and Snell method shows, R^2 value of the fitted model is 0.055 less than 1, which say that the model is perfect. According to the Nagelkerke method, R^2 value is 0.075, which is between 0 and 1 interpret as a good model.

The Hosmer and Lemshow test is used to check the adequacy of the fitted model and the results are shown in Table 5.

H_0: The model adequately fit the data
H_1: The model does not adequately fit the data

Table 5: Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	.000	0	.
2	.986	2	.611
3	3.475	4	.482
4	7.435	4	.115
5	4.280	5	.510
6	5.280	6	**.508**

According to Table 5, there is significant evidence that the model adequately fit the data.

CONCLUSION

The Fitted model is,

$$\text{Logit}(P_{ijklmn}) = (\text{Sector})_i + (\text{Nature})_j + (\text{Location})_k + (\text{Gender})_l + (\text{No. of employee})_m + (\text{Legal status})_n$$

Where, P_{ijklmn} is the probability of a changing of SLSIC code at i^{th} sector, j^{th} nature, k^{th} location, l^{th} gender, m^{th} no. of employees and n^{th} legal status.

This model concludes that, there are six variables affecting to the change of SLSIC code. Those are sector, nature, location, gender, No. of employees and Legal status.

The best way to control non-sampling errors is to follow the right procedures of all survey activities from planning sample selection up to the analysis.

REFERENCES

- Assael, H., & Keon, J. (1982, April). Nonsampling vs. Sampling Errors in Survey Research. *Journal of Marketing, 46*(2), 114.
- Bond, D. (1992). Documentation of Nonsampling Error Issues in the Survey of Industrial Research and Development.
- Jeremiah P., B. (2003, December). Nonsampling errors in surveys. *United Nations Statistics Division.*
- Roberts, C. (2007). *Mixing modes of data collection in surveys:A methodological review.* London: ESRC National Centre for Research Methods Briefing Paper.