Using calibration in a Survey on Transportation of Goods by Road

Juris Breidaks[1]

[1]Central Statistical Bureau of Latvia, e-mail: Juris.Breidaks@csb.gov.lv

Abstract

The paper is devoted to the quality analysis of the ongoing survey “Transportation and Turnover of Goods by Road” organised by the Central Statistical Bureau of Latvia (CSB). This is a continuous survey. The stratified simple random sampling is used.

The summary estimation of the variance is compared with the summary GREG estimation of the variance in the paper. These estimations are compared by the sample error.

1 Introduction

The Survey on Transport of Goods by Road was initiated in January 1997 as a pilot project organized by *Eurostat* under the *Phare Programme*. It is a continuous survey where information about the vehicles in the sample is obtained through questionnaires mailed to respondents. The target of survey is to obtain the information about transportation of goods by road performed by transport vehicles registered in Latvia. The main variables of interest are tonnes transported, tonne-kilometres performed and kilometres travelled loaded for total goods road transport.

The survey covers transport vehicles that are owned by legal and natural persons and which at the moment of sample formation had undergone technical inspection and could be lawfully used. The data of the Road Traffic Safety Directorate about vehicle registrations and the number of vehicles that had undergone technical inspection reveal, and could be legally used. Special vehicles such as fire-fighting engines, crane lorries, tower cranes, road repair vehicles and other special vehicles were not included in the survey.

Simple random stratified sampling is used. The weekly sample size is 100 vehicles.

2 Stratification

For 2010 stratification is made by capacity, place of registration of vehicles, year of release of the vehicles, status.

Table 1 – Stratification for 2010

Stratum	Capacity and place of registration of vehicles	Year of release of the vehicles	Status of person
3	3,5t<cap. ≤ 5t, Riga(including the district of Riga)	All	Legal
4	3,5t<cap. ≤ 5t, all Latvia without Riga and the district of Riga	All	Legal
5	5t<cap. ≤10t, Riga(including the district of Riga)	2005-2011	Legal
6	5t<cap. ≤ 10t, Riga(including the district of Riga)	1998-2004	Legal
7	5t<cap. ≤ 10t, Riga(including the district of Riga)	1991-1997	Legal
8	5t<cap. ≤ 10t, all Latvia without Riga and the district of Riga	2005-2011	Legal
9	5t<cap. ≤ 10t, all Latvia without Riga and the district of Riga	1998-2004	Legal
10	5t<cap. ≤ 10t, all Latvia without Riga and the district of Riga	1991-1997	Legal
11	cap.>10t, Riga(including the district of Riga)	2005-2011	Legal
12	cap.>10t, Riga(including the district of Riga)	1998-2004	Legal
13	cap.>10t, Riga(including the district of Riga)	1991-1997	Legal
14	cap.>10t, all Latvia without Riga and the district of Riga	2005-2011	Legal
15	cap.>10t, all Latvia without Riga and the district of Riga	1998-2004	Legal
16	cap.>10t, all Latvia without Riga and the district of Riga	1991-1997	Legal
17	the trucks, Riga(including the district of Riga)	2005-2011	Legal
18	the trucks, Riga(including the district of Riga)	1998-2004	Legal
19	the trucks, Riga(including the district of Riga)	1991-1997	Legal
20	the trucks, all Latvia without Riga and the district of Riga	2005-2011	Legal
21	the trucks, all Latvia without Riga and the district of Riga	1998-2004	Legal
22	the trucks, all Latvia without Riga and the district of Riga	1991-1997	Legal
31	3,5t<cap. ≤ 5t, all Latvia	All	Private
32	5t<cap., all Latvia	All	Private
33	the trucks, all Latvia	All	Private

3 The Horvitz – Thomson estimator and variance

Unit design weights are calculated according to the sampling design and inclusion probabilities of units in the sample – $w_{dh} = \frac{N_h}{n_h}$, where N_h is population size of stratum h and n_h is the sample size in stratum h.

Final weight of unit i are calculated as $w_h = w_{dh} \cdot \frac{n_h}{n_h^r} = \frac{N_h}{n_h} \cdot \frac{n_h}{n_h^r} = \frac{N_h}{n_h^r}$, where N_h is population size of stratum h and n_h is the number of respondents in stratum h.

The Horvitz – Thomson estimator (HT) estimator is

$$\hat{Y}_{HT} = \sum_{i=1}^{n^R} y_i w_i$$

and its estimated variance is

$$\hat{V}(\hat{Y}) = \sum_{h=1}^{H} \frac{1 - \frac{n_h^R}{N_h}}{n_h^R} \frac{1}{n_h^R - 1} \sum_{i=1}^{n_h^R} \left(w_i y_i - \frac{1}{n_h^R} \sum_{i=1}^{n_h^R} w_i y_i \right)^2$$

where n^R – number of respondents in sample;

H – number of strata in sampling frame;

y_i – value of study variable of unit i ; $y = (y_1, \ldots, y_n)$

w_i – final weight of unit i . $w = (w_1, \ldots, w_n)$

4 GREG estimator and variance

Set of responded transport vehicles in each month is assumed to be a sample. New frame in each month is assumed as population of transport vehicles in beginning of the month.

In each month sample was calibrated on the new frame. The number of respondents in each strata and the capacity of vehicles (legal persons - 3,5t<capacity ≤ 5t, 5t < capacity ≤10t, capacity>10t, the trucks, private persons – total capacity.) has been used as auxiliary variables.

Package “sampling” of software *R* is used for the calibration, and g-weights are calculated with the help of a function “calib” from this package. Whereas calibration is based on the “raking” method in the function “calib”.

Please observe, that using GREG (Generalized Regression) estimator (calibration), the weights are not equivalent in one stratum within.

The GREG estimator is

$$\hat{Y}_{GREG} = \sum_{i=1}^{n^R} y_i w_i g_i$$

and its estimated variance is

$$\hat{V}_{GREG}(\hat{Y}) = \sum_{h=1}^{H} \frac{1 - \frac{n_h^R}{N_h}}{n_h^R} \frac{1}{n_h^R - 1} \sum_{i=1}^{n_h^R} \left(w_i g_i e_i - \frac{1}{n_h^R} \sum_{i=1}^{n_h^R} w_i g_i e_i \right)^2$$

where estimated residual is

$$\hat{e} = y - X_s \cdot \left((X_s \cdot w)^T \cdot X_s \right)^{-1} (X_s \cdot w)^T \cdot y$$

5 Results

The variance of estimates was estimated for 15 indicators – total tonnes transported, total tonne-kilometres performed and total kilometres travelled loaded for total goods road transport and for national goods road transport, and for export goods road transport, and for import goods road transport, and for international goods road transport.

Table 2 – The coefficients of variation for estimates of indicators in year 2010

	Quarter							
	1		2		3		4	
	HT	GREG	HT	GREG	HT	GREG	HT	GREG
TONN	8,01	7,94	8,02	7,19	8,96	8,46	9,60	9,22
TKM	4,46	4,46	4,69	4,74	4,30	4,29	4,95	4,98
KML2	3,80	3,72	3,75	3,73	3,50	3,44	3,89	3,82
TO_N	10,18	10,12	9,57	8,47	10,48	9,90	11,38	10,99
TK_N	8,76	8,79	8,43	8,32	7,70	7,68	7,76	7,72
KM_N	7,19	7,05	6,61	6,24	5,92	5,79	5,96	5,76
TO_EXP	10,22	10,02	15,90	18,32	9,00	9,13	10,58	10,41
TK_EXP	7,92	7,85	7,64	7,68	8,10	8,17	9,22	9,44
KM_EXP	7,52	7,37	6,87	6,84	7,22	7,19	8,53	8,60
TO_IMP	15,95	16,59	18,19	19,01	18,10	17,55	12,38	12,21
TK_IMP	13,43	13,38	12,06	12,12	12,01	11,87	13,08	12,99
KM_IMP	9,14	9,04	8,17	8,20	8,13	7,99	8,32	8,27
TO_INT	11,35	11,25	12,70	12,37	12,91	12,85	12,95	12,90
TK_INT	10,54	10,54	11,18	11,37	10,49	10,40	11,90	11,97
KM_INT	9,37	9,39	9,85	10,07	9,62	9,57	10,39	10,35

Notations for this table

TONN Tonnes transported for total goods road transport

TKM Tonne-kilometres performed for total goods road transport

KML2 Kilometres travelled loaded for total goods road transport

TO_N Tonnes transported for national goods road transport

TK_N Tonne-kilometres performed for national goods road transport

KM_N Kilometres travelled loaded for national road transport

TO_EXP Tonnes transported for export goods road transport

TK_EXP Tonne-kilometres performed for export goods road transport

KM_EXP Kilometres travelled loaded for export goods road transport

TO_IMP Tonnes transported for import goods road transport

TK_IMP Tonne-kilometres performed for import goods road transport

KM_IMP Kilometres travelled loaded for import national road transport

TO_INT Tonnes transported for international goods road transport

TK_INT Tonne-kilometres performed for international goods road transport

KM_INT Kilometres travelled loaded for total international road transport

References

Särndal C.-E., Sundström S., “Estimation in the presence of nonresponse and frame imperfections, Statistics Sweden”

Särndal C.-E., Sundström S., (2005) Estimation in Surveys with Nonresponse.Wiley, England.

Särndal C.-E., Swensson B., Wretman J. (1992) *Model Assisted Survey Sampling.* Springer-Verlag, New York.

Lapins, J. (1997) Sampling surveys in Latvia: Current situation, problems and future developments. *Statistics in Transition: Journal of the Polish Statistical Association*, 3, 281-292.

Pandutang V, Sukhatme, Balkrishna V. Sukhatme , Shashikala Sukhatme, C. Asok., Sampling Theory of Surveys with Applications, IOWA State University press and Indian Society of Agricultural Statistics, 14-15

Lohr S. L. (1999) *Sampling: Design and Analysis.* Brooks/Cole Publishing Company, Pacific Grove, Calif.

Guillaume OSIER The linearisation approach implemented by EUROSTAT for the first wave of EU-SILC: What could be done from the second wave onwards? Luxembourg Income Study (LIS), Institut National de la Statistique et des Etudes Economiques (STATEC Luxembourg) and Net-SILC2. Paper prepared for the workshop on standard error estimation and other related sampling issues (Eurostat, 29-30 March 2012)

SPSS Inc. (2012) *SPSS® Syntax Reference Guide*

Liberts. M. (2004) *Quality Analysis of a Sample Survey on Transportation of Goods by Road*.