Using calibration in a Survey on Transportation of Goods by Road

Juris Breidaks
Central Statistical Bureau of Latvia
E-mail: Juris.Breidaks@csb.gov.lv

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

- 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

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.

Stratification

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

Table 1 – Stratification for 2011

The Horvitz – Thomson (HT) estimator and variance

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

- y_i – value of study variable of unit i
- w_i – weight of unit i
- n^R – number of respondents
- N_h – population size of strata h
- **Variance estimator**

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

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.

As auxiliary variables has been used

- The number of respondents in each strata
- The capacity of vehicles
  - ✓ legal persons – 3,5t<capacity <= 5t, 5t < capacity<=10t, capacity>10t, the trucks
  - ✓ private persons – total capacity.

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".

The GREG estimator is $\hat{Y}_{GREG} = \sum_{i=1}^{H} y_i w_i g_i$

And its estimated variance is

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

where estimated residual is

$$\hat{e}_i = y_i - X_s \times \left((X_s \times w_i)^T \times X_s\right)^{-1} \times (X_s \times w_i) \times y_i$$

Results

	Quarter							
	1		2		3		4	
	HT	GREG	HT	GREG	HT	GREG	HT	GREG
TONN	8,0	7,9	8,0	7,2	9,0	8,5	9,6	9,2
TKM	4,5	4,5	4,7	4,7	4,3	4,3	4,9	5,0
KM2	3,8	3,7	3,8	3,7	3,5	3,4	3,9	3,8
TO_N	10,2	10,1	9,6	8,5	10,5	9,9	11,4	11,0
TK_N	8,8	8,8	8,4	8,3	7,7	7,7	7,8	7,7
KM_N	7,2	7,1	6,6	6,2	5,9	5,8	6,0	5,8
TO_EXP	10,2	10,0	15,9	18,3	9,0	9,1	10,6	10,4
TK_EXP	7,9	7,8	7,6	7,7	8,1	8,2	9,2	9,4
KM_EXP	7,5	7,4	6,9	6,8	7,2	7,2	8,5	8,6
TO_IMP	16,0	16,6	18,2	19,0	18,1	17,6	12,4	12,2
TK_IMP	13,4	13,4	12,1	12,1	12,0	11,9	13,1	13,0
KM_IMP	9,1	9,0	8,2	8,2	8,1	8,0	8,3	8,3
TO_INT	11,3	11,2	12,7	12,4	12,9	12,9	13,0	12,9
TK_INT	10,5	10,5	11,2	11,4	10,5	10,4	11,9	12,0
KM_INT	9,4	9,4	9,8	10,1	9,6	9,6	10,4	10,4

Table 2 – The coefficients of variation (CV) for estimates of indicators in year 2011

Conclusion

- Do not get the desired result that CV for estimates of indicators in year 2011 with GREG is not given better result as CV for estimates of indicators in year 2011 with HT.
- Need to search auxilary variables which better describe data.