QUALITY ANALYSES OF LABOUR FORCE SURVEY 2007

Andris Fisenko[1], Jānis Jukāms[2]

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

[2] Central Statistical Bureau of Latvia,
University of Latvia
e-mail: janis.jukams@csb.gov.lv

Abstract

The main goal of the research is to compare sampling errors of Labour Force Survey (LFS) 2007 with the forecasts of sampling errors produced in 2006 and make analysis of monthly estimates. Starting from January 2007 sample size for LFS survey has been increased. This abstract contains main results on the monthly estimates of the number of employed and unemployed persons.

1. Introduction

From January 2007 LFS total sample size has been increased almost 2.4 times. The goal of this step was to ensure production of the monthly, quarterly and annual employment and unemployment statistics of the population (persons aged 15 – 74 years) and in different domains of interest (by region, sex, different age groups, etc.). In order to improve the precision of the estimates some administrative data was combined with the survey data. As one of such sources the Register of unemployed persons of the State Employment Agency (SEA) was used in the process of analysis.

2. Description

The stratified two-stage probability sampling principles are used in the LFS survey. The sample allocation between strata is made proportional to the population sizes of strata. LFS stratified by the degree of urbanisation. Counting areas of population census are used as the primary sampling units. At the second stage simple random sampling is used. The dwellings/households register is used as a sampling frame. The number of population from household register is used as calibration variables. These calibration variables are divided in 112 groups (4 strata, sex, 14 age groups). Also number of population in 6 regions is used. SEA database contains person level information on registration of persons in that database as unemployed persons this information also can be used for calculation of g - weights and implementation of the generalized regression estimators for production of the monthly LFS estimates has been made.

We present generalised regression (GREG) estimator as

$$\hat{Y}_{GREG} = \hat{Y} + (X - \hat{X})'B, \tag{1}$$

where

$\hat{Y}$ - *Horvitz-Thompson* estimate of the total for study variable.

X - the vector of known auxiliary totals

$\hat{X}$ - the vector of *Horvitz-Thompson* estimates of the totals for auxiliary variables

B - the vector of regression coefficients obtained by linear fitting of Y on X, using the data belonging to the sample. It is estimated by:

$$\hat{B} = \left(\sum_{i=1}^{n} q_i \frac{x_i x_i'}{\pi_i} \right)^{-1} \left(\sum_{i=1}^{n} q_i \frac{x_i y_i}{\pi_i} \right) \tag{2}$$

q_i in formula (2) stands for known distance weights (usually equal to 1, otherwise take on values in the range [0..1], π_i - inclusion probabilities and n – number of respondents. The term $(X - \hat{X})'B$ in the formula (1) can be viewed as the regression adjustment to the Horvitz-Thompson estimator.

We can also explain *GREG* estimator as

$$\sum_{i=1}^{n} \frac{y_i g_i}{\pi_i} = \sum_{i=1}^{n} y_i w_i ,$$

where w_i are calibrated weights derived as

$$w_i = \frac{1}{\pi_i} g_i$$

and g-weights expressed in the form:

$$g_i = 1 + (X - \hat{X})\hat{T}^{-1} q_i x_i ,$$

having $\hat{T}$ equal to

$$\hat{T} = \left(\sum_{i=1}^{n} q_i \frac{x_i x_i'}{\pi_i} \right).$$

The exact variance formula has a complex form, so we used an approximation of it, expressed by the residuals of the regression. The residuals are estimated as follows:

$$\hat{e}_i = y_i - x_i'\hat{B}$$

And the Jackknife linearization variance of the GREG estimator:

$$\hat{V}_{JL}\left(\hat{V}_{GREG}\right)=\sum_{h}\left(1-\frac{n_h}{N_h}\right)\frac{1}{n_n\left(n_h-1\right)}\sum_{j=1}^{n_h}v_{hj}^2$$

where h – index of strata, j – index of cluster inside stratum h, n_h - number of clusters in the sample of stratum h and N_h - number of clusters in the population of stratum h. v_{hj} is an empirical influence value for the cluster hj.

As the *GREG* estimator has been used for level estimation v_{hj} is computed as

$$v_{hj}=n_h e'_{hj}-\sum_i e'_{hj}\ ,$$

having $e'_{hj}=\sum_i w_{hji}e_{hji}$ with w_{hji} - calibrated weight, e_{hji} -regression residual and i – index for the unit inside cluster j.

For ratio estimation v_{hj} is computed as

$$v_{hj}=\frac{v_{hj}^Y-\hat{R}v_{hj}^Z}{\hat{Z}},$$

having v_{hj}^Y - empirical influence value for the Y variable, v_{hj}^Z - empirical influence value for the Z variable, $\hat{R}$ - estimated value for the ratio of $\hat{Y}$ to estimated value for the Z level.

3. Results

The main country-level parameters for which it is planned to produce the monthly LFS estimates are total number of employed persons, total number of unemployed persons, employment rate and unemployment rate. In addition, estimates of these parameters have to be produced by sex and two age groups (15-24, 25-74), too. Based on LFS 2006 data twelve monthly estimates of all parameters previously mentioned and their sampling errors were calculated see Table1.

Table 1 The annual average of the estimated coefficient of variation, in percent, for monthly estimates of LFS 2007 data (Forecast estimates based on LFS 2006 data)

	Number of employed persons		Number of unemployed persons		Employment rate		Unemployment rate	
	Forecast	2007	Forecast	2007	Forecast	2007	Forecast	2007
Latvia, all persons 15-74	1.9	1.6	11.1	11.2	1.9	1.5	11.2	10.9
Males, 15-24	7.9	7.6	31.1	32.2	7.9	7.4	31.6	31.9
Males, 25-74	2.3	1.9	16.1	16.9	2.3	1.9	16.3	16.5

Females, 15-24	11.0	10.8	33.0	39.9	11.0	10.6	33.9	38.3
Females, 25-74	2.5	2.2	17.5	18.0	2.5	2.2	17.5	17.6

Estimates for number of employed and employed rate derived from data of the year 2007 exceed previously obtained forecasts produced in 2006, but unemployed and unemployed rate forecasts exceed the estimates derived from LFS 2007 data. See chart 2 and chart 3.

Most of results for monthly average LFS estimates are satisfying, but with some exceptions. The estimated coefficient of variation of the total number of unemployed in the age group 15-24 is approximately equal to 32% and for females in the same group is approximately equal to 39%. All results are average numbers from monthly data and for (these) two groups Table 2 shows how big the differences are between the coefficients of variation in one month.

Table 2. The minimum and maximum of the estimated coefficient of variation, in percent, for estimates of LFS 2007- (number of unemployed persons).

	Min		Max	
Males, 15-24	24.1	(June)	56.0	(October)
Females, 15-24	28.1	(June)	57.0	(September)

Since the sample size within domains of 15-24 years old males and females is too small this can cause the big coefficient of variation. There is a connection between the minimum and the maximum values of the coefficients of variation and month. For example the minimum is reached in June when lot of students and pupils start summer working. The maximum is reached when this group of people stop working.

Chart 2. Difference between coefficients of variation employed persons.

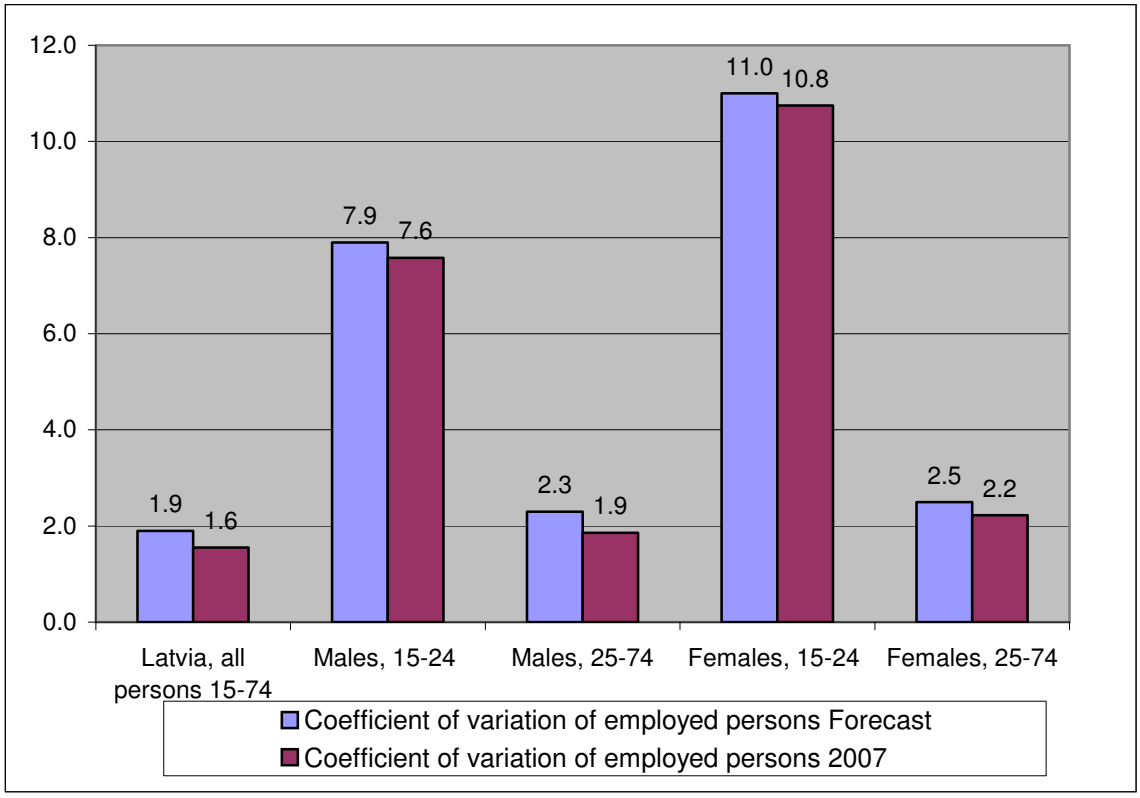
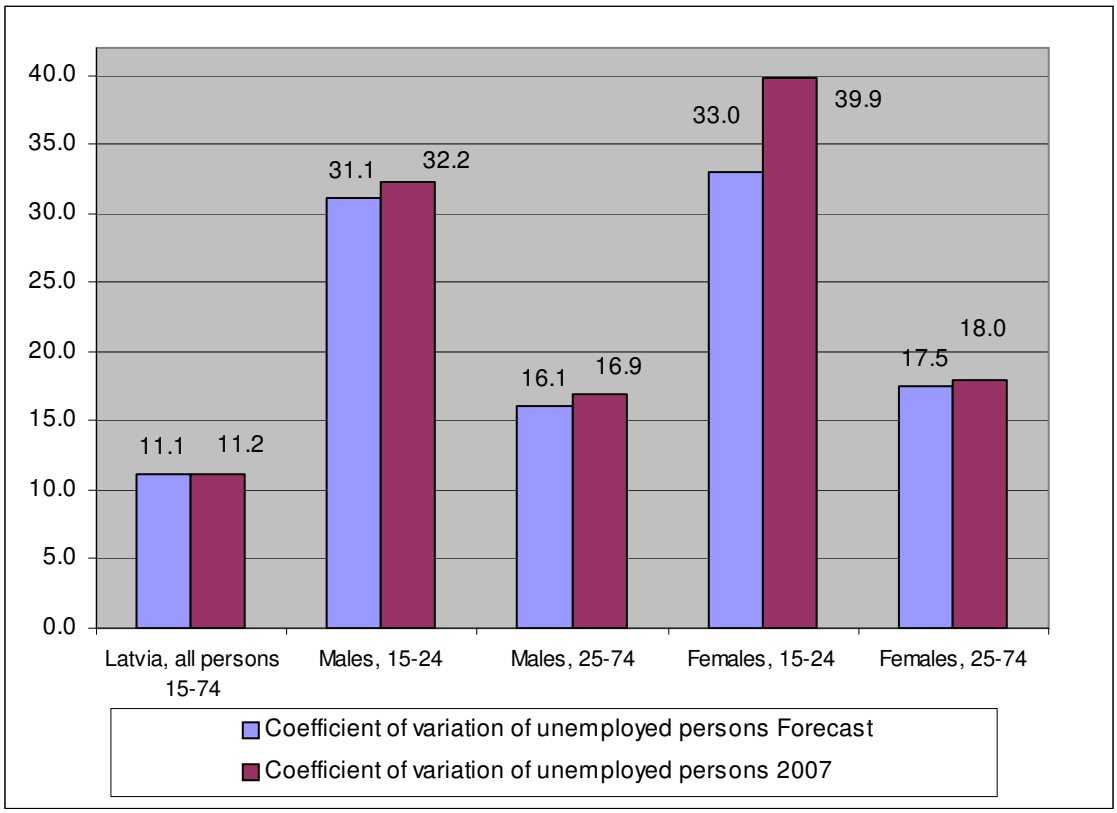


Chart 3. Difference between coefficients of variation unemployed persons.



4. Conclusion

These are only the first results from analysis of LFS data 2007. To achieve our main goal – possibility to publish monthly results - a lot of work still needs to be done. Recommendations for future activities:

- Try to get faster monthly estimations, not only after end of quarter;
- Calculate coefficient of variation for monthly estimations in some different groups (regions, districts and etc.)
- To combine LFS data with data from Register of unemployed persons of the SEA at person level and to produce generalized regression estimates. To calculate sampling errors of these estimates.
- For forecasting monthly estimates of LFS indicators to corresponding time series in order to offer possibilities to analyse data using time series analysis tools.

References

- Ollila, P. (2004) A theoretical Overview for Variance Estimation in Sampling Theory with Some New Techniques for Complex Estimators, Research Reports, No. 240, Statistics Finland.
- Wolter, K.M. (1985) Introduction to Variance Estimation, Springer Verlag.
- Lundström S., Särndal C.-E. (2005) Estimation in Surveys with nonresponse. Wiley, Chichester.