Notater
Documents
46/2013

Statistics Norway
Statistisk sentralbyrå

Jørn Ivar Hamre and Johan Heldal

Improved calculation and dissemination of coefficients of variation in the Norwegian LFS

Documents 46/2013

Jørn Ivar Hamre and Johan Heldal

Improved calculation and dissemination of coefficients of variation in the Norwegian LFS

Statistisk sentralbyrå • Statistics Norway
Oslo–Kongsvinger

Documents In this series, documentation, method descriptions, model descriptions and standards are published.

Published November 2013

ISBN 978-82-537-8791-6 (printed)
ISBN 978-82-537-8792-3 (electronic)
ISSN 1891-5906
Subject: Labour market and earnings

Print: Statistics Norway

Symbols in tables	**Symbol**
Category not applicable	.
Data not available	..
Data not yet available	…
Not for publication	:
Nil	-
Less than 0.5 of unit employed	0
Less than 0.05 of unit employed	0.0
Provisional or preliminary figure	*
Break in the homogeneity of a vertical series	—
Break in the homogeneity of a horizontal series	|
Decimal punctuation mark	.

Preface

This report documents the implementation of improved calculation and dissemination of coefficients of variation in the Norwegian Labour Force Survey. It also summarizes studies related to non-sampling errors in the Norwegian LFS.

The document is mainly the final report to Eurostat from one of two projects at Statistics Norway granted by European Commission in 2011 for "Quality improvement actions for LFS".

The report has been written by Jørn Ivar Hamre at Division for labour market statistics and Johan Heldal at Division for statistical methods.

Statistisk sentralbyrå, 6 September 2013

Hans Henrik Scheel

Abstract

The estimation procedure in the Norwegian Labour Force Survey (NLFS) is more complicated than simple calibration, but an empirical variance estimator was derived for the NLFS by Hagesæther and Zhang (Notater 2007/22) based on linearization.

This report documents the implementation of the improved calculation of sampling error for the main variables and groups in the regular production system of the NLFS. The empirical variance estimator is also extended with covariance elements to cover figures of change and annual averages. The new calculations of coefficient of variation (CV) and standard error (SE) are published in connection with the regular publication of quarterly and annually NLFS figures in the StatBank on our webpage. The quarterly CV and SE figures are published in the StatBank table http://www.ssb.no/en/table/09937 and annual figures in http://www.ssb.no/en/table/09938.

Some of the results are presented and discussed in this report. Average of quarterly standard error in 2011 and 2012 for total unemployment and total employment are 5.1 and 0.33 per cent of the estimated values respectively.

Eurostat has proposed new precision requirements. One of the proposed requirements is about the standard error for the estimated annual average of the proportion of unemployed at NUTS 2 level (region). Our calculations show that we will fulfil this proposed requirement for all the 7 Norwegian regions in 2010, 2011 and 2012.

The other proposed precision requirement, is about the standard error for the difference in estimated proportion of unemployed at national level between two successive quarters. In spite of the high overlap between samples in adjacent quarters, the NLFS would not fulfil this new proposed precision requirement for change estimates of unemployed persons. One reason for this is the low autocorrelation for unemployment, so the high overlap of samples is of little help for making good estimates of change in the unemployment. However, the high overlap of sample makes the change estimates for employment better. Also, the NLFS estimation procedure does not include any good register predictors for LFS-unemployment, but is in stead optimized for making good quarterly county-divided employment figures.

Also other sources of survey errors in the NLFS are described in this report.

Contents

Preface **3**

Abstract **4**

Contents **5**

1. Introduction **6**

2. Project description in our grant application **6**

3. Data and the methods of the Norwegian LFS **8**
3.1. Sampling in the Norwegian LFS 8
3.2. Weighting 9
3.3. Calculation of the variance with linearization 10

4. Some results of the new standard error and coefficient of variation calculations **13**
4.1. Developments in SE and CV in the Norwegian LFS since 2006 14
4.2. Do the Norwegian LFS fulfil the new proposed precision requirements? 16
4.3. Figures of CV in Accuracy Reports to EUROSTAT for 2012 19

5. Non-sampling errors **20**
5.1. Frame and design issues 20
5.2. Response issues 20
5.3. Measurement issues 21
5.4. Processing issues 22

6. Fulfilment of the project goals **23**

References **24**

List of figures **25**

List of tables **25**

1. Introduction

This document is the final report from the project II: "Improved calculation and dissemination of coefficients of variation in the Norwegian LFS" according to the EUROSTAT GRANT AGREEMENT No. 10201.2011.001-2012.184

This is a documentation of the actions of this project and gives a description of some results.

The following people been working in the project: Johan Heldal (Division for statistical methods), Jørn Ivar Hamre (Division for labour market statistics) and Solveig Lyby, (Division for statistics systems)

Chapter 3.3 of this report is written by Johan Heldal, while Jørn Ivar Hamre has written the other chapters.

Chapter 2 is the project description. It describes the objective, actions and the results that were the goals for the project. Chapter 3 describes our data and methods in the Norwegian Labour Force Survey (NLFS). Some of the results of the new Standard Error (SE) and Coefficient of Variation (CV) calculations are presented in chapter 4, where we focus on main trends, and whether we fulfil the new proposed precision requirement for the LFS. We also compare the new calculations with the old for the quarterly accuracy reports to Eurostat. Different aspects of non-sampling errors in the NLFS are described in chapter 5. In chapter 6 we evaluate the goals of the project.

2. Project description in our grant application

Objective:

The goal of this project is to implement the calculation of sampling error in the regular production system of the Norwegian Labour Force Survey (LFS), and to publish the figures every quarter for the main variables and groups. This will also improve the "Quarterly accuracy report" we send to Eurostat.

The project aims at better fulfilling recommendation 29 of the Task force on the quality of the LFS, which is "Accompany published estimates by information about their accuracy, covering in principle both sampling and non-sampling errors".

Situation:

After the major changes in the Norwegian LFS in 1996, the official figures for sampling error or standard error available to users[1] have not been updated.
In 1996, the rotation of the samples was changed, mainly to reduce the statistical uncertainty for estimates on short-term changes. On the other hand, this change led to an increased standard error for annual averages. Now each family participates in the survey eight times during a period of eight quarters. 33 000 different individuals are now interviewed annually, compared to 54 000 before this change.

The preliminary figures of coefficient of variation (CV) in our yearly quality reports to EUROSTAT are not based on our actual method of estimation, and have not been published at our website to all the users.

Possible non-sampling errors are to some extent described at our web-page "About the statistics". N**on-sampling errors** can be divided into the following four broad groups 1) **Frame** and **design** issues, 2) **Non-response** issues, 3) **Measurement issues and 4) Processing** issues.

[1] web-page "About the statistics" (www.ssb.no/aku_en/about)

Background
The estimation procedure in the Norwegian LFS is more complicated than simple calibration.

The weights (inflation factors) are adjusted in three steps with calibration on counties, register employment, age groups and gender after post-stratification nationally on register employment, age groups and gender. The initial weights take account of different sampling and non-response proportions in each county. Our sampling design is a one-stage county-stratified systematic sampling (with assumed random sorting) of ***family clusters*** from our Central Population Register.

In Hagesæther and Zhang (2009) the design effect on the estimation uncertainty of family clustering is analysed in the situation of a calibration estimator with good auxiliary variables. They show that the use of auxiliary information removes the extra variance that is due to the variation in the cluster sizes. Therefore, Hagesæther and Zhang (2007) derived an empirical variance estimator for the Norwegian LFS where they, for simplicity, assume that persons are sampled instead of families. The estimator is based on linearization and some asymptotical assumptions (Zhang, 2006), and is easy to implement.

The empirical variance estimators for the *quarterly* weights used before the changes in our LFS in 2006 have already been programmed in SAS. However it has been challenging to implement the programs and integrate them with our regular production system. This report describes this implementation. In our regular production system we compute *monthly* weights. The *quarterly* weights are the *monthly* weights divided by 3. After the changes in our LFS in 2006, age is defined per the 15th of the months and therefore makes the delimitation of population files different each month.

Description of action
1. Prepare the quarterly samples for the calculation of sampling error
2. Implement the calculation of sampling error in the regular production system of the Norwegian LFS.
3. Design and calculate tables of SE and CV needed for the accuracy report to EUROSTAT and for checking of the precision requirement in the Regulation
4. Start publishing them quarterly StatBank
5. Document and update our web-page “About the statistics” with more detailed information about each of the four issues described as non-sampling errors
6. Refine the “Quarterly accuracy report” to EUROSTAT with the new figures.
7. Documentation and reporting to EUROSTAT

We plan to publish the following tables in our StatBank:
SE and CV for employed persons by age groups and gender. Quarterly and annually
SE and CV for the number of part-time employed persons by age groups and gender. Quarterly and annualy
SE and CV for the number of unemployed persons by age groups and gender. Quarterly and annually
SE and CV for the rate of unemployment by age groups and gender. Quarterly and annually
SE and CV for the average actual hours of work per week by age groups and gender. Quarterly and annually
SE and CV for employed persons by county and region and gender. Annual average.
SE for the change in employed persons from previous quarter by age groups and gender.
SE for the change in unemployed persons from previous quarter by age groups and gender.

SE for the change in employed persons from the same quarter last year by age groups and gender.
SE for the change in unemployed persons from the same quarter last year by age groups and gender.

3. Data and the methods of the Norwegian LFS

3.1. Sampling in the Norwegian LFS

Our sampling design is a one-stage county-stratified systematic sampling (with assumed random sorting) of family clusters from our Central Population Register, where lesser populated counties are overrepresented. The sample consists of about 12 000 family units (24 000 persons) each quarter.

Sampling frame

The sampling frame is based on the Central population register (CPR). The CPR is updated daily, with an 8 days deadline. The Norwegian Tax Administration owns the CPR, and Statistics Norway use quarterly data files from the CPR both for sampling and for estimation.

Sampling unit

Family is the primary sampling unit in the Norwegian LFS. The unit consists of family[2] members aged 15 – 74 years[3] in the resident population according to the CPR.

The current definition of family includes the following types of families:
single persons, married couples without children, married couples and their children and single parent who live with child(ren).

The sampling frame consists of registered family units where the main person in the family is aged 15-74 years.

A detailed discussion about the actual sampling unit ***family*** in relation the ***household*** is given in Villund (2010).

A note about collective households

Persons living in some collective households may be surveyed through their family. Student for instance, living in a hostel at a University or College, may be surveyed through family if they are not registered as moved from their parents in the CPR, which is the general rule. Persons enrolled in compulsory military service or living in other institutions can also be surveyed, but not people registered as living in prison or mental institution. In general, if persons are registered as moved from their parent(s) they are their own family and can be sampled and surveyed independent of whether they are living in a private or collective household.

Stratification and allocation of the sample

The sample is stratified by the 19 counties (NUTS3 regions) in Norway. This means that we divide the population of families after the county of residence to the main person in the family, and thereafter sample families from each county separately.

[2] The family concept in the Norwegian CPR is not the same as the family concept used by ESS and Eurostat since it includes individuals living alone. According to the family definition in the CPR a family can consists of members from maximum 2 generations. A reference person is defined (the male in the parent generation, possibly a female if no male exists) in every family, and the unique personal ID number to the reference person is used as the family ID number.

[3] Also 13 and 14 year olds in sampled families are included, but put on hold until they turn 15.

Lesser populated counties are disproportionately overrepresented in order to make the coefficient of variation (for employment figures) more even between different counties. The table under shows the over/under sampling rate in different counties, which have been constant over time.

Table 3.1. Relationship between sampling proportions in different counties

County	Over/under-sampling ratio
Oslo, Akershus, Hordaland	0.871
Aust-Agder	1.484
Sogn og Fjordane	1.398
Nord-Trøndelag	1.151
Finnmark	1.851
Other counties	1

Source: Vedø and Rafat (2003)

The quarterly sample of about 12 000 family units (24 000 persons) is distributed evenly among the 13 reference weeks in the quarter. In other words, the NLFS is a continuous survey. A month in LFS consists of either 4 or 5 whole reference weeks, so the monthly gross sample consists of about 24000*4/13 or 24000*5/13 persons.

Sampling method

For each county, family units are sampled systematically from a list of all the families in the population in the county. The sorting of the lists is by municipality, postal code, street number, house number and family-ID-number. In the estimation the sorting of the lists is assumed to be random within each county.

Rotation pattern

From 1996 onwards each sampled family participates in the survey 8 times every 13 week during a two year period. This means that 7/8 of the total sample overlap between two adjacent quarters, and that half of the total sample overlap between a quarter and the same quarter a year before[4].

Updating of the sample

One eighth of the sample is renewed every quarter. All the persons aged 15-74 originally sampled are followed in the two year period, independent of what happens with family relations (divorce, marriage) or relocations within Norway of families or family members. No persons are added to the sample other than originally sampled 13 and 14 year olds, who are put on hold until they are 15 years. No persons are removed from the sample[5] in the two year period other than people turning 75 year, deceased and emigrants.

Data collection in the Norwegian LFS

All resident persons aged 15 – 74 years in the selected families are interviewed by telephone (computer assisted). Proxy interviewing is to some extent allowed if direct interview is not possible.

3.2. Weighting

We compute *monthly* weights based on monthly data sets. The *quarterly* weights in the *quarterly* data files are the *monthly* weights divided by 3. The estimation

[4] According to Steel & McLaren (2008) this rotation pattern is good for the following estimates: 3 months average and change between adjacent 3 months periods, 3-months average of seasonally adjusted data and change in them and level and change in trend estimates. On the other hand, our rotation pattern is poor for making change estimates between months 1 and 2 months apart. Also the very high overlap rate in Norway gives fewer independent observations in a year than many other rotation patterns, and therefore gives less precise estimates of annual averages.

[5] Also persons with the following non-response reasons are not interviewed again, but still remains in the gross sample: a) Long-lasting sick (new contact out of question) and b) Persistent refusal in the follow-up.

procedure in the Norwegian LFS is more complicated than simple calibration. The weights (inflation factors) are adjusted three times with calibration on counties, register employment, 12 age groups and gender after national post-stratification on 3 groups of register employment, 12 age groups and gender. The initial weights take account of different sampling and non-response proportions in each county.

In the step of county calibration we use a bisected register status: registered non-employed (RNE) and register employed (RE), while in step of the national post-stratification RE is divided in 3 industry groups, except for persons 70 – 74 years. The industry groups are 1) Primary industry, 2) Manufacturing and construction, and 3) Service industry.

About Register information

There are different updating and delay of registers.

Quarterly files from the CPR define resident population at the end of the months in the middle of the quarter. Age is defined per the 15th of the respective months, and therefore makes the delimitation of (age group divided) population files different each month. When preliminary weights for the first months in the quarter is made (for input to seasonal adjustment) the quarterly file from the CPR is not ready yet, so we use CPR from the previous months.

We also utilize auxiliary information from Register of Employees and the Registers of unemployed at the Employment office, which both are owned by the Norwegian Labour and Welfare Service (NAV).

The Norwegian annual tax Register (from two year earlier) are also utilized to identify self employed persons, which is a part of the register employed. (The Register of Employees is the main input for auxiliary variable register employment in the population file.)

For the time being[6], the register of unemployed at the Employment office is only used for a consistency check between the registers. The register of unemployed dominate the employee register and the annual tax register, so registered unemployed at the Employment office are classified as RNE together with everybody else that is not RE.

As from 2011, we started to use monthly "just in time" produced information from the Register of Employees and the Register of unemployed at the Employment office at the end of each months. Before 2011 we only used Register of Employees and the Register of unemployed at the Employment office for the middle months of the quarter for estimation of all the months in the quarter.

Detailed documentation of the method of estimation in the Norwegian Labour Force Survey (NLFS) is given in Zhang (1998, 2000) and Heldal (2000).

3.3. Calculation of the variance with linearization

An empirical variance estimator are derived for the Norwegian LFS that are easy to implement based on linearization in Hagesæther and Zhang (2007).

Assumption

Hagesæther and Zhang (2007) do not take account of clusters in the empirical variance estimator. Due to small differences, they assume that we sample persons instead of families for simplicity. In Hagesæther and Zhang (2009) the design

[6] Earlier findings from micro linking of the register of unemployed at the Employment office and the NLFS sample indicates that agreement rate between LFS-unemployed and registered unemployed is not that high, especially not for young and old people.

effect on uncertainty estimation of family clustering is analysed in the situation of a calibration estimator with good auxiliary variables[7]. It is shown that the use of auxiliary information removes the extra variance that is due to the variation in the cluster sizes. Therefore we are assuming simple random sampling within each county.

Let s be the entire sample, P be either the entire national target population for the LFS or any subpopulation defined by county or the calibration variables age and gender and $sP = s \cap P$, the part of the sample that falls into subpopulation P. Let W_i be the final calibrated weight for individual no. i in the sample s. The total T_Y for a variable Y in the subpopulation P is estimated as

$$\hat{T}_{YP} = \sum_{i \in sP} W_i Y_i$$

Let n_P be the sample size falling in subpopulation P. Let e_{Yi} be the residuals for individual no. i from the fit of the generalized regression model behind the county-wise calibration. The calibration weighted residuals ($W_i e_{Yi}$) are the key elements for all estimation of variances and covariances demonstrated in this section. Let $\overline{We}_{YP} = \sum_{i \in sP} W_i e_{Yi} / n_P$ be the mean weighted residual in sP. The variance of $\hat{T}_Y$ can be estimated by:

$$\hat{V}_{YP} = \frac{n_P}{n_P - 1} \sum_{i \in sP} \left(W_i e_{Yi} - \frac{\sum_{j \in s} W_i e_{Yi}}{n_P} \right)^2 = \frac{n_P}{n_P - 1} \left(\sum_{i \in sP} \left(W_i e_{Yi} \right)^2 - n_P \overline{We}_{YP}^2 \right)$$

There are different residuals for different variables. Behind the above equation are a few asymptotical assumptions which will not be presented here (Zhang, 2006). The standard error of $\hat{T}_{YP}$ is estimated as $\widehat{SE}(\hat{T}_{YP}) = \sqrt{\hat{V}_{YP}}$. The estimated coefficient of variation (CV) is the estimated standard error in per cent of the estimated value, $\widehat{CV}(\hat{T}_{YP}) = 100 \widehat{SE}(\hat{T}_{YP}) / \hat{T}_{YP}$.

Covariance over time and between variables

For estimation of variances for changes, ratios (e.g. unemployed in per cent of the labour force) and annual averages, covariance estimates over time and between different variables must be calculated. The covariance between the totals $\hat{T}_{YP}$ and $\hat{T}_{ZP}$ in the same quarter are estimated by

$$\hat{C}(\hat{T}_{YP}, \hat{T}_{ZP}) = \frac{n_P}{n_P - 1} \left(\sum_{i \in sP} W_i^2 e_{Yi} e_{Zi} - n_P \overline{We}_Y \overline{We}_Z \right)$$

In order to estimate variances for changes in single variables or in ratios, covariances between totals for subpopulations at different quarters are needed. Let the subscripts 1 and 2 refer to first and last quarter respectively and subscripts 12 refer to the sample overlap between the two quarters. This overlap will generate correlations between estimates for the two quarters. Since the sample each quarter consists of eight waves, of which only one is replaced each quarter, there will be sample overlaps up to seven quarters apart. Let $sP1$ and $sP2$ mean the sample for the subpopulation P for the two quarters and $sP12 = SP1 \cap sP2$, the overlap between the two samples. Their sample sizes are denoted n_{P1}, n_{P2} and n_{P12}

[7] NLFS is used as an example in chapter 4 of the article.

respectively. The covariance between the estimates $\hat{T}_{YP1}$ for T_{YP} at quarter 1 and $\hat{T}_{ZP2}$ for T_{ZP} at quarter 2 are then estimated by

$$\hat{C}(\hat{T}_{YP1},\hat{T}_{ZP2}) = \frac{n_{P12}}{n_{P12}-1}(\sum_{i\in sP12} W_{i1}e_{Yi1}W_{i2}e_{Zi2} - n_{P12}\overline{W_1e_{Y1}}\,\overline{W_2e_{Z2}}).$$

Note that the variable *Y* at quarter 1 and *Z* at quarter 2 may well be the same variable for (e.g. employed/unemployed for quarters 1 and 2 respectively).

Having produced the estimates of the variances and covariances shown above, the variances of ratios and of the estimates of changes are produced from these estimates. Let *a* and *b* refer to quarter for the estimates so that *a*, *b* = 1 or 2. We may have *a* = *b* which means the same quarter.
The variance of the ratio $\hat{R} = \hat{T}_{YPb} / \hat{T}_{ZPa}$ is estimated by

$$\hat{V}(\hat{R}) = (\hat{V}(\hat{T}_{YPb}) + \hat{R}^2\hat{V}(\hat{T}_{ZPa}) - 2\hat{R}\hat{C}(\hat{T}_{YPb},\hat{T}_{ZPa}))/T^2_{ZPa}$$

The covariance between estimates for two ratios, which is needed for estimation of variances of estimates of change in e.g. the unemployment rate, can be estimated as

$$\begin{aligned}\hat{C}(\hat{R}_1,\hat{R}_2) &= \hat{C}(\frac{\hat{T}_{Y1}}{\hat{T}_{Z1}},\frac{\hat{T}_{Y2}}{\hat{T}_{Z2}}) \\ &= \left(\hat{C}(\hat{T}_{Y1},\hat{T}_{Y2}) - \hat{R}_1\hat{C}(\hat{T}_{Z1},\hat{T}_{Y2}) - \hat{R}_2\hat{C}(\hat{T}_{Y1},\hat{T}_{Z2}) + \hat{R}_1\hat{R}_2\hat{C}(\hat{T}_{Z1},\hat{T}_{Z2})\right)/\hat{T}_{Z1}\hat{T}_{Z2}\end{aligned}$$

In cases where we want ratios measured in percentages, such as the percentage of unemployed of the work force, $\hat{R}$ is multiplied by 100 and their variances and standard errors likewise. The variance of a simple change $\hat{D}_{YP12} = \hat{T}_{YP2} - \hat{T}_{YP1}$ from quarter 1 to quarter 2 is estimated by

$$\hat{V}(\hat{D}_{YP12}) = \hat{V}(\hat{T}_{YP1}) + \hat{V}(\hat{T}_{YP2}) - 2\hat{C}(\hat{T}_{YP1},\hat{T}_{YP2})$$

The relative change

$$\widehat{RD}_{YP12} = \frac{\hat{D}_{YP12}}{\hat{T}_{YP1}} = \frac{\hat{T}_{YP2}-\hat{T}_{YP1}}{\hat{T}_{YP1}} = \frac{\hat{T}_{YP2}}{\hat{T}_{YP1}} - 1$$

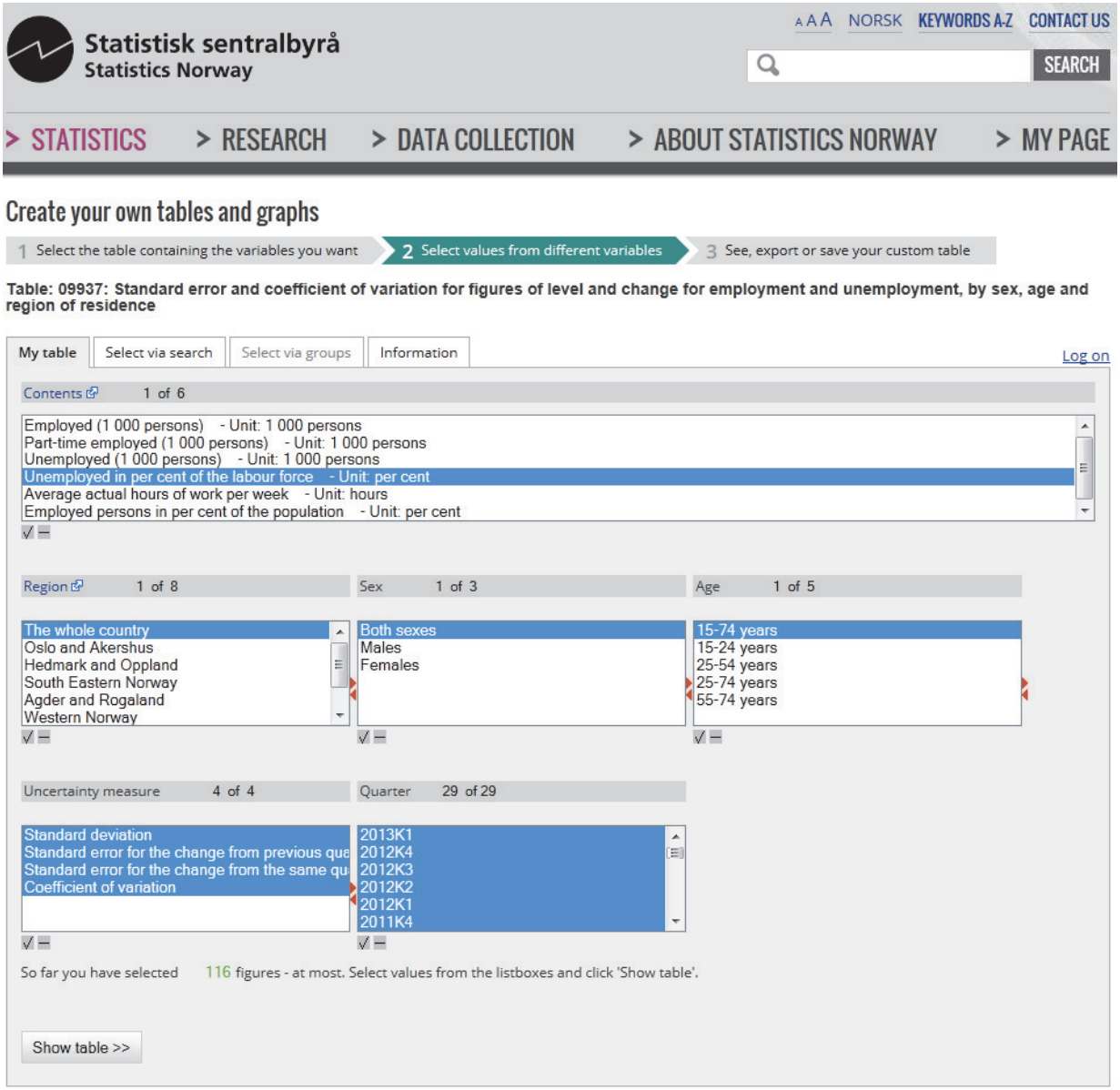
and its variance is just the variance of the ratio $\hat{T}_{YP2} / \hat{T}_{YP1}$.

For a variance estimate of an annual average, the covariance estimate between periods 1, 2 and 3 quarters apart are utilized. Let q = 1, 2, 3, 4 denote the four quarters within a year. The yearly average for a total is estimated as $\hat{T}_{YPyear} = \sum_{q=1}^{4}\hat{T}_{YPq}/4$. Its variance is estimated as

$$\hat{V}(\hat{T}_{YPyear}) = (\sum_{q=1}^{4}\hat{V}(\hat{T}_{YPq}) + 2\sum_{q=1}^{3}\sum_{t=q+1}^{4}\hat{C}(\hat{T}_{YPq},\hat{T}_{YPt}))/16.$$

4. Some results of the new standard error and coefficient of variation calculations

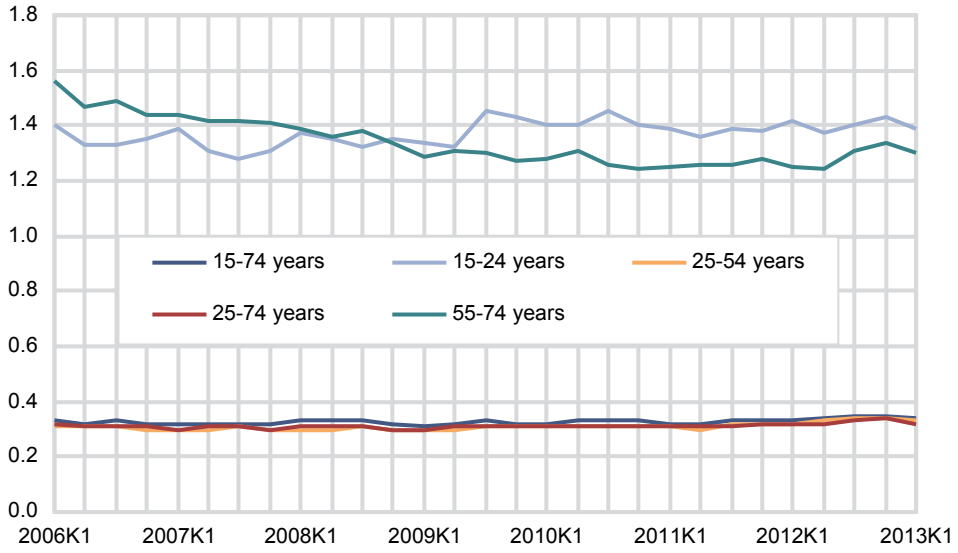
Our new calculations of coefficient of variation (CV) and standard error (SE) will be published in connection with the regular publication of quarterly and annually LFS figures on the webpage to Statistics Norway. The SE and CV calculations for quarterly figures are published in the StatBank tables: http://www.ssb.no/en/table/09937 and http://www.ssb.no/en/table/09938 for annually figures.



From our StatBank the user can create his or her own tables and graphs by selecting table, and selecting values from different variables.

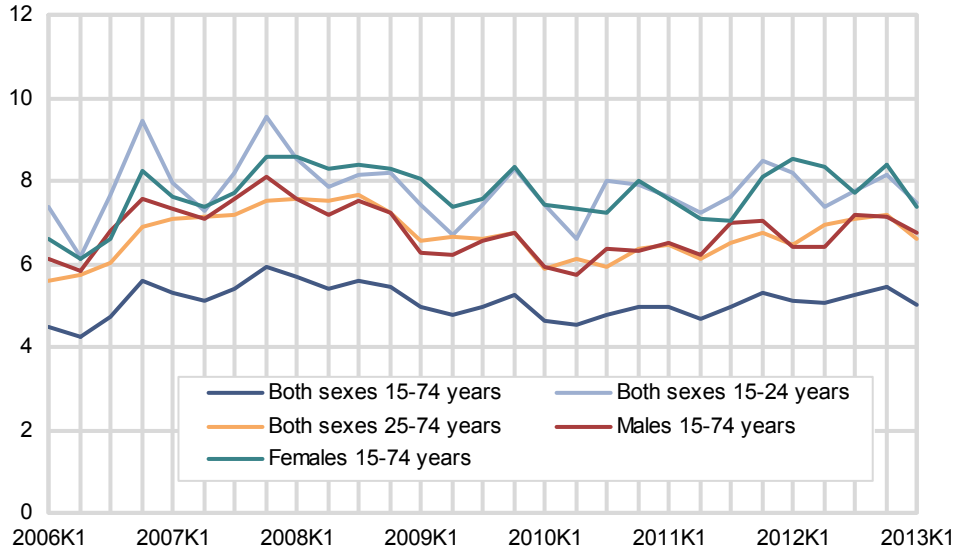
4.1. Developments in SE and CV in the Norwegian LFS since 2006

Figure 4.1. Coefficient of variation for employed persons by age



Coefficient of variation for employed persons is stable around 0.3 per cent for the total and for the age group 25-54 and 25-74. For the age group 15-24, the coefficient of variation have been around 1.4 per cent since 3rd quarter 2009.

Figure 4.2. Coefficient of variation for unemployed persons by age and sex

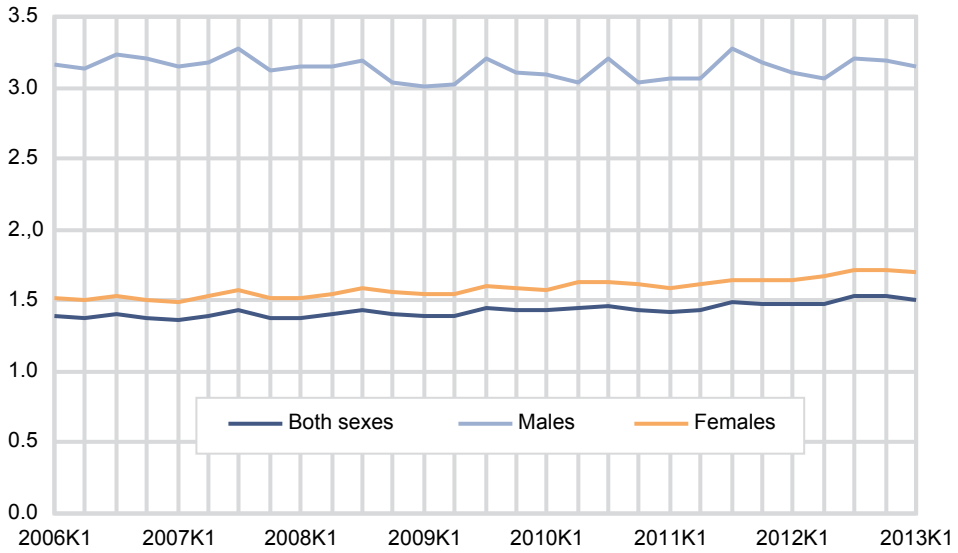


The new calculations of coefficient of variation for quarterly averages of unemployed persons in the NLFS have been around to 5 per cent since 2010, and around 5.5 percent in 2007 and 2008.

Mainly due to seasonal variations in the denominator, the coefficient of variation for quarterly averages of unemployed persons 15-24 years roughly has been between 7 and 8 per cent since 2008. The unemployment figures for persons 25-74 years are little bit more precise, but the CV have gradually increased from around 6% in the beginning of 2010 to about 7% at the end of 2012. The CV for female

unemployment has been between 1 and 2 percentage points higher than for men since 2008.

Figure 4.3. Coefficient of variation for part-time employed persons aged 15-74, by sex



The coefficient of variation for part-time employed is on the upper edge of 3 per cent for females 15-24 and around 2 per cent for females 25-54. For the total of males and females the CV is around 1.5 per cent.

Figure 4.4. Coefficient of variationfor part-time employed persons, by age and sex

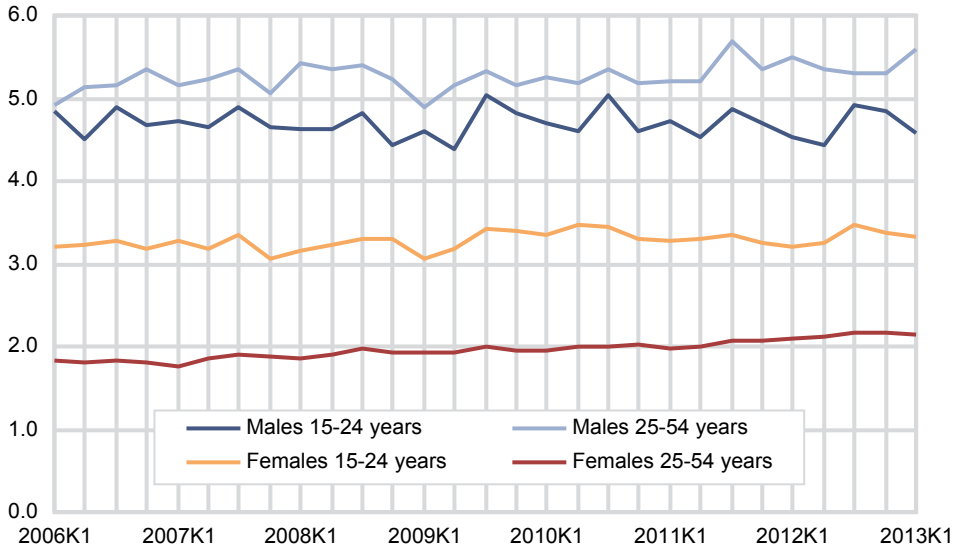
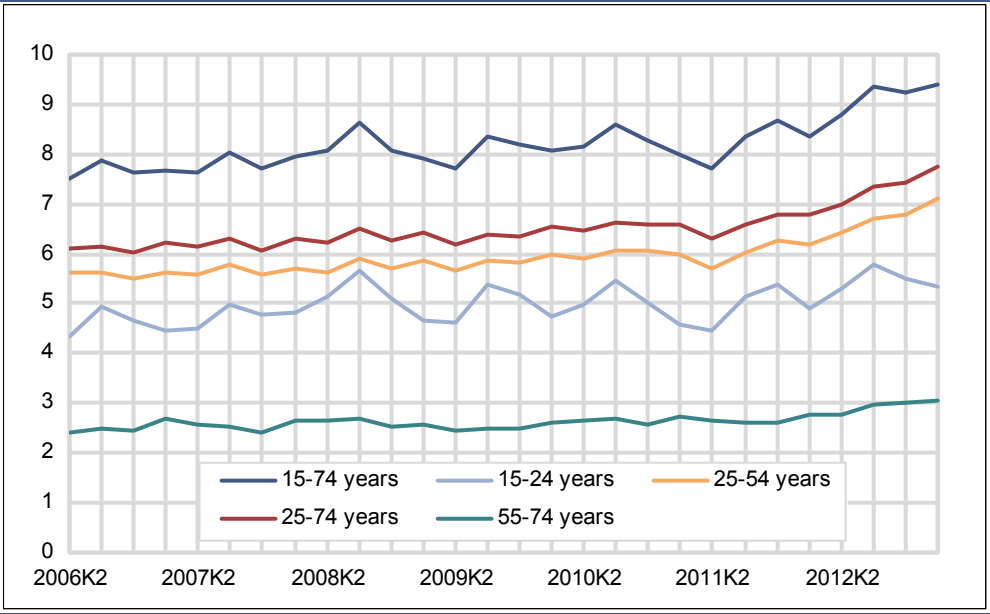
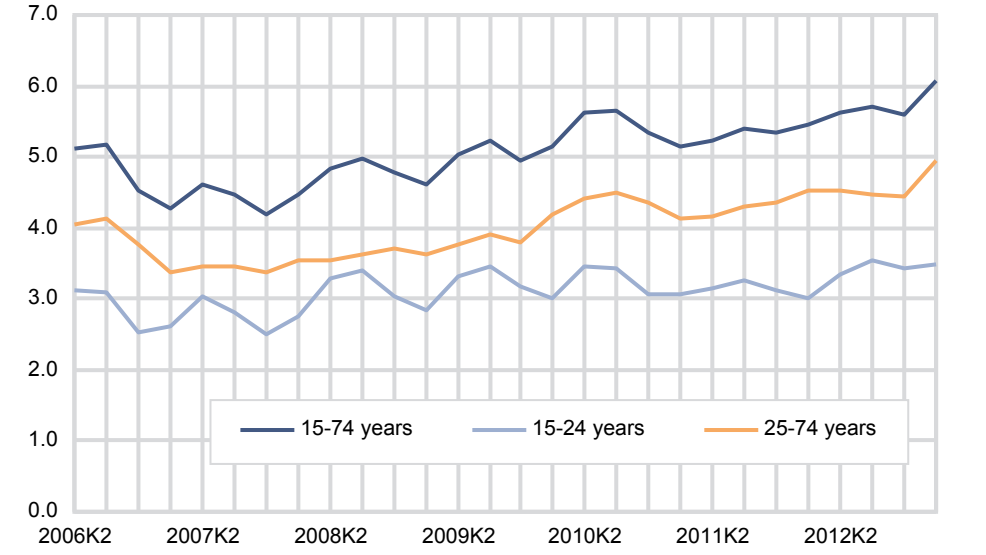


Figure 4.5. Standard error for the change from previous quarter in the number of employed persons, by age groups (in 1 000)



Standard error (S.E.) for the change from previous quarter in the number of employed persons seems to have increased from 2nd quarter 2011 for the total. Behind this, is especially an increase in the S.E. for the quarterly change in employment for the age group 25-54, which increased with about 1400 persons.

Figure 4.6 Standard error for the change from previous quarter in the number of unemployed persons, by age groups (in 1 000)



The standard error (S.E.) for the change from previous quarter in the number of unemployed persons seems to have increased gradually for the total from about 4 thousand in the beginning of 2007 to about 6 thousand in 1st quarter 2013.

4.2. Do the Norwegian LFS fulfil the new proposed precision requirements?

In the documents[8] from the group of experts on precision requirements for the LFS two new precision requirement was proposed.

[8] Doc. Eurostat/F2/EMPL/16/10 and doc. Eurostat/F2/LAMAS/38/10

Requirement for a revised article 3.1

At NUTS 2 level, the standard error for the estimated annual average of the proportion of unemployed to the population aged 15 to 74, $SE(\hat{p}_t)$, shall not exceed $\sqrt{\hat{p}_t * (1-\hat{p}_t)/2969}$, with exemption for the NUTS 2 regions with less than 300,000 inhabitants.
In this check we use the formula $SE(\hat{p}_t) = SE(U_t / N_t) = SE(U_t)/N_t$, where U_t is the annual average of the estimated value of unemployed persons in a Norwegian region, and N_t is the annual average the quarterly population figure in a Norwegian region from the CPR, so it is not a stochastic variable. $SE(U_t)$ is already calculated in http://www.ssb.no/en/table/09938.

Table 4.1. Standard error for the estimated annual average of the proportion of unemployed to the population aged 15 to 74, by year and region

Year	Region	$SE(\hat{p}_t)$	$\sqrt{\hat{p}_t * (1-\hat{p}_t)/2969}$	$\hat{p}_t$
2010	NO01-Oslo and Akershus	0.0017	0.0032	0.0314
2010	NO02-Hedmark and Oppland	0.0023	0.0027	0.0223
2010	NO03-Sør-Østlandet	0.0016	0.0029	0.0260
2010	NO04-Agder and Rogaland	0.0015	0.0025	0.0196
2010	NO05-Vestlandet	0.0016	0.0029	0.0249
2010	NO06-Trøndelag	0.0021	0.0029	0.0263
2010	NO07-Nord-Norge	0.0020	0.0029	0.0257
2011	NO01-Oslo and Akershus	0.0016	0.0029	0.0262
2011	NO02-Hedmark and Oppland	0.0021	0.0025	0.0194
2011	NO03-Sør-Østlandet	0.0015	0.0028	0.0244
2011	NO04-Agder and Rogaland	0.0013	0.0024	0.0169
2011	NO05-Vestlandet	0.0016	0.0028	0.0236
2011	NO06-Trøndelag	0.0022	0.0029	0.0257
2011	NO07-Nord-Norge	0.0019	0.0028	0.0246
2012	NO01-Oslo and Akershus	0.0015	0.0028	0.0246
2012	NO02-Hedmark and Oppland	0.0022	0.0026	0.0200
2012	NO03-Sør-Østlandet	0.0016	0.0029	0.0255
2012	NO04-Agder and Rogaland	0.0016	0.0025	0.0194
2012	NO05-Vestlandet	0.0015	0.0027	0.0216
2012	NO06-Trøndelag	0.0022	0.0028	0.0245
2012	NO07-Nord-Norge	0.0019	0.0027	0.0228

According to our calculation the standard error for the estimated annual average of the proportion of unemployed to the population aged 15 to 74, $SE(\hat{p}_t)$, is less than $\sqrt{\hat{p}_t * (1-\hat{p}_t)/2969}$ for all the 7 Norwegian regions in 2010, 2011 and 2012, and therefore fulfil one of the new proposed precision requirements.

Requirement for a revised art. 3.2

At national level, the standard error for the difference in estimated proportion of unemployed to the population aged 15 to 74 between two successive quarters, $SE(\hat{p}_{t+1} - \hat{p}_t)$, shall not exceed $\sqrt{\hat{\bar{p}} * (1-\hat{\bar{p}})/21111}$ for countries with a population between one and twenty millions, which is the case fore Norway[9].

For simplicity, we assume that quarterly population figure is unchanged between two successive quarters $N_{t+1} = N_t$. For this check, we then use:

[9] $\hat{\bar{p}}$ could be the value of the estimate either at time *t*, or at time *t+1*, or the average between the two values, or the max between the two.

$$SE(\hat{p}_{t+1} - \hat{p}_t) = SE\left(\frac{U_{t+1}}{N_{t+1}} - \frac{U_t}{N_t}\right) = SE\left(\frac{U_{t+1} - U_t}{N_{t+1}}\right) = \frac{SE(U_{t+1} - U_t)}{N_{t+1}}$$

where U_t is the of the estimated quarterly average of unemployed persons at national level.

$SE(U_{t+1} - U_t)$ is already calculated in http://www.ssb.no/en/table/09937.

Table 4.2. Standard error for the difference in the estimated proportion of unemployment to the population aged 15 to 74 between two successive quarters

Periode (t)	$SE(\hat{p}_t - \hat{p}_{t-1})$	$\sqrt{\hat{p}*(1-\hat{p})/21111}$	$\hat{p}_t$
2006K2	0.0015	0.0012	0.029
2006K3	0.0015	0.0011	0.024
2006K4	0.0013	0.0009	0.018
2007K1	0.0012	0.0009	0.019
2007K2	0.0013	0.0010	0.020
2007K3	0.0013	0.0009	0.018
2007K4	0.0012	0.0009	0.016
2008K1	0.0013	0.0009	0.019
2008K2	0.0014	0.0010	0.021
2008K3	0.0014	0.0009	0.019
2008K4	0.0014	0.0009	0.019
2009K1	0.0013	0.0010	0.022
2009K2	0.0014	0.0011	0.025
2009K3	0.0015	0.0010	0.024
2009K4	0.0014	0.0010	0.021
2010K1	0.0014	0.0011	0.026
2010K2	0.0016	0.0011	0.028
2010K3	0.0016	0.0011	0.025
2010K4	0.0015	0.0010	0.023
2011K1	0.0014	0.0010	0.023
2011K2	0.0014	0.0011	0.026
2011K3	0.0015	0.0010	0.023
2011K4	0.0014	0.0010	0.022
2012K1	0.0015	0.0010	0.023
2012K2	0.0015	0.0011	0.024
2012K3	0.0015	0.0010	0.022
2012K4	0.0015	0.0010	0.023
2013K1	0.0016	0.0011	0.026

The above results of standard error for the difference in the estimated proportion of unemployed to the population aged 15 to 74 between two successive quarters shows that the NLFS would not have fulfilled the new proposed precision requirement since $SE(\hat{p}_t - \hat{p}_{t-1}) > \sqrt{\hat{p}*(1-\hat{p})/21111}$ for all time periods.

In spite of the high overlap between adjacent quarters, the NLFS would not have fulfilled the new proposed precision requirement for change estimates of unemployed persons. One reason for this is the low autocorrelation for unemployment, so the high overlap is of little help for making good estimates of change in the unemployment. Also, the NLFS estimation procedure does not include any (good) register predictors for LFS-unemployment, but is in stead optimized for making good quarterly county-divided employment figures. According to Hagesæther and Zhang (2009) the estimated variances of the GREG estimators are slightly increased using the register information, which is drawback for using good auxiliary information for employment that at the same time is ineffective for unemployment. High association between the auxiliary information and response propensity in combination with low association between auxiliary information and a LFS target variable leads to variance inflation for the LFS target variable according to Zhang, Thomsen and Kleven (2013).

Also, sample size and allocation and sampling methods are determinants of the precision[10].

[10] See Thomsen and Zhang (2001)

4.3. Figures of CV in Accuracy Reports to EUROSTAT for 2012

Table 4.3. Coefficient of variation (CV) for quarterly estimates. 2012

Period	Employed persons	Part-time employed persons	Unemployed persons	Unemployed in per cent of the labour force	Average actual hours of work per week
2012K1	0.33	1.47	5.11	5.03	0.36
2012K2	0.34	1.48	5.08	5.00	0.39
2012K3	0.35	1.53	5.26	5.18	0.38
2012K4	0.35	1.53	5.43	5.34	0.38
2012 average	0.21	0.98	2.91	2.87	0.22

Comparing these new 4th quarter CV estimates with earlier sent figures in the Quarterly Accuracy report, shows that for employed persons the new CV is 0.03 percentage points lower, the new CV for part-time employed is 0.17 percentage points higher. For unemployed persons, the new CV is 0.05 percentage points lower. Also the CV estimate or the average actual hours of work per week[11] is 0.15 percentage points lower.

Table 4.4. Coefficient of variation (CV) for annual estimates, by region. 2012

Region	Employed persons	Part-time employed persons	Unemployed persons	Unemployed in per cent of the labour force	Average actual hours of work per week
Norway	0.21	0.98	2.91	2.87	0.22
Oslo and Akershus	0.46	2.31	6.25	6.16	0.43
Hedmark and Oppland	0.78	3.37	11.09	10.92	0.85
Sør-Østlandet	0.51	2.21	6.34	6.24	0.52
Agder and Rogaland	0.52	2.32	8.46	8.36	0.54
Vestlandet	0.51	2.35	7.00	6.91	0.54
Trøndelag	0.69	3.15	8.94	8.79	0.73
Nord-Norge	0.66	3.17	8.46	8.33	0.72

Table 4.5. Coefficient of variation (CV) of annual estimates by county (NUTS3). 2012

County	NUTS3-code	Coefficient of variation (CV)			
		Number of persons in the labour force	Number of employed	Number of unemployed	Unemployment rate
01 Østfold	NO031	0.97	1.03	10.8	10.7
02 Akershus	NO012	0.64	0.66	9.4	9.3
03 Oslo	NO011	0.62	0.64	8.4	8.2
04 Hedmark	NO021	1.08	1.09	15.7	15.5
05 Oppland	NO022	1.12	1.13	15.6	15.4
06 Buskerud	NO032	0.92	0.94	14.8	14.6
07 Vestfold	NO033	0.95	0.98	12.5	12.3
08 Telemark	NO034	1.12	1.14	13.2	12.9
09 Aust-Agder	NO041	1.43	1.44	20.4	20.1
10 Vest-Agder	NO042	1.08	1.10	16.4	16.2
11 Rogaland	NO043	0.63	0.64	11.0	10.9
12 Hordaland	NO051	0.66	0.68	10.0	9.9
14 Sogn og Fjordane	NO052	1.20	1.22	21.4	21.2
15 Møre og Romsdal	NO053	0.95	0.97	10.9	10.7
16 Sør-Trøndelag	NO061	0.81	0.83	10.4	10.2
17 Nord-Trøndelag	NO062	1.18	1.21	17.2	17.0
18 Nordland	NO071	0.89	0.91	12.8	12.6
19 Troms Romsa	NO072	1.16	1.20	15.7	15.4
20 Finnmark Finnmárku	NO073	1.37	1.44	14.3	14,1

[11] The new CV are calculated for the sum of actual hours worked in 1st and 2nd jobs, and restricted to those who actually worked 1 hour or more in the reference week.

5. Non-sampling errors

Non-sampling errors can be clustered in the following four broad groups 1) **Frame** and **design** issues, 2) **Non-response** issues, 3) **Measurement issues and 4) Processing** issues.

Possible non-sampling errors are to some extent described at our web-page "About the statistics".
Here we try to summarize information for NLFS on issues of non-sampling errors that have been studied at Statistics Norway.

5.1. Frame and design issues

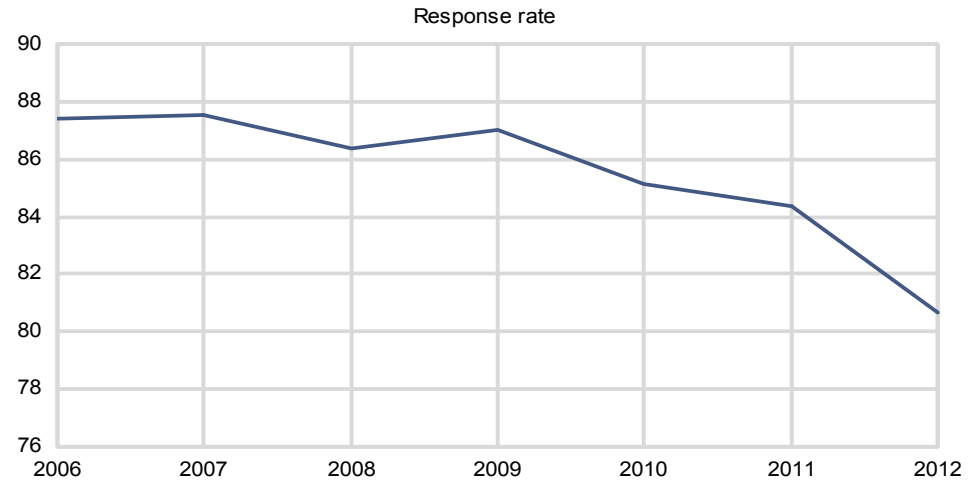
Preliminary findings by the Division for statistical methods suggest that the gross sample of the NLFS may not be as representative as we used to think. They linked the gross sample with registered employment divided by industry, and the results suggests that register employed systematically are a little over represented in the gross sample. They also find that the gross sample gives a higher proportion register employed in the major industry division Human health and social work activities than in the whole register. The weighting procedure adjust for this over representation of registered employed, but don't seem to correct all the detailed industry divided figures that well. This may be a design effect due to sampling of families or the county-stratified sampling where counties have different sampling fractions. In this new preliminary study we did not have the exact same population register that we selected the sample from, and so far we can not rule out the possibility that some of this effect is due to the removal of persons living in prison and mental institution from the sampling register.

The sampling frame consists of registered family units where the main person in the family is aged 15-74 years. Due to the fact that on average women marry older men, old women are under sampled because some of them have a husband over 74 years and is therefore out of today's specified sampling frame.

5.2. Response issues

The non-response rate was approximately 10-12 per cent in the years 1972-1991, and for the next six years just 6-8 per cent. From 1998 onwards the non-response in the NLFS has increased gradually, up to approximately 20 per cent in 2012.

Figure 5.1. Yearly average of quarterly response rate in the Norwegian LFS



In the NLFS there is more non-response among registered non-employed (RNE) than among register employed (RE). Our estimation procedure seems to correct for much of that by post-stratification on RE and RNE among other things.

Trends in non-response and its effects in NLFS is analysed in Thomsen and Villund (2012) using administrative information linked to gross sample through a unique Personal Identification Number (PIN). They have estimated the nonresponse bias by using register-based employment status available for the gross sample for different subgroups over time. They find that the bias of registered employment is more or less constant in the NLFS, across several subgroups such as age, gender and nationality, during a period with increased nonresponse rates. The correlation between LFS employment and register employment is very high[12], so these findings are relevant for the LFS variable as well. The result suggests that nonresponse rate alone is not a good indicator of quality[13].

For other variables non-response bias seems to be a major concern. Thomsen and Villund have also made comparisons (not yet published) of differences in the registered unemployment rate in net and gross samples of NLFS linked through a unique PIN to register data over time. The results indicate increasing non-response bias for the registered unemployment rate over last couple of years, when we have seen a rising non-response rate.

Using data from the NLFS and administrative registers, Thomsen and Zhang (2001) demonstrate that the use of registers in has little or no additional effect on the accuracy of estimates of change based on the **panel part** of the survey data, neither in terms of the sampling variance nor in the bias introduced by non-response. The main reason for this is that the administrative registers available are not sufficiently up-to-date at the time of production. After 2010, the NLFS utilize more up-to-date monthly registers in the estimation procedure. Hopefully this procedure makes the short-term change figures more accurate when the economic cycles are changing.

Zhang (2005) evaluate and compare the bias due to non-response and misclassification in the sample quarter-to-quarter **gross labour flow** estimates between the labour force statuses (employed, unemployed and not in the labour force) in the NLFS. His main conclusions are the following: (a) the overall labour market stability, i.e., the proportion of people without change in status, should be boosted after adjusting for both non-response and misclassification, (b) neither non-response nor misclassification affects the net change estimates, and (c) misclassification has very little effect on the level estimation of the characteristics “employed”, “unemployed” and “not in the labour force”.

5.3. Measurement issues

Due to cost consideration, some proxy interviewing is allowed in the NLFS, and about 15 percent of the total sample is proxy interviews. For some subgroups, especially young people, the proxy rate is much higher. For youth proxy interviewing is made through their parents. Lack of correct information may lead to measurement errors or item non-response. Measurement error because of proxy interviewing probably varies between different variables. Due to a lot of proxy interviewing about young people, there is a substantial underestimation of employment for young people (Villund, 2010).

Proxy interviews provide data on some hard-to-reach people who have a labour-market situation more similar to that of those not reached at all. Therefore proxy

[12] See Thomsen and Zhang (2001).

[13] They define response structure as a ratio of response rates, and a simple relationship between the bias of a binomial variable and the response structure is presented. From this relationship they see that the bias is a function of the response structure and not of the nonresponse rate.

interviews probably result in a better total employment rate estimate, even though they introduce some underreporting. That is the main result from a study using register data to evaluate the effects of proxy interviews in the Norwegian LFS (Thomsen and Villund, 2011)[14]

For variables such as actual hours of work and (hours of) overtime measurement errors may very well be a major concern. For the overtime and underemployment variables data by proxy interviewing are rejected in the NLFS, and treated as partial non-response in stead. Separate overtime figures are not yet published, and a method of imputation is developed.

5.4. Processing issues

Computer-assisted interviews are done by telephone. Information from previous interview is used in order to save time. For the coding of industry, information from some registers is also used. Data on education are based on a register of individual data collected by Statistics Norway from the educational institutions (but questions are also asked to get more updated information).

As all interviews are computer-assisted, some procedures for electronic control of the registration of answers are included in the questionnaire, for example concerning the number of working hours during the reference week. In some cases the interviewers get a "warning" when recording an answer, in other cases maximum or minimum values have been set beforehand.

Processing issues are also discussed in Zhang, Thomsen and Kleven (2013) using standard process indicators for fieldwork monitoring in a cross-survey perspective, where LFS is one of many surveys at Statistics Norway.

[14] Also see Zhang, Thomsen and Kleven (2013) for an extension of the approach.

6. Fulfilment of the project goals

Table 4.6. Table of action and fulfilment

No.	Action	Fulfilled	Comments
1.	Prepare the population files for the calculation of sampling error	YES	
2.	Implement the calculation of sampling error in the regular production system of the Norwegian LFS.	YES	
3.	Design and program tables of CV needed for the accuracy report to EUROSTAT and for checking of the precision requirement in the Regulation	YES	See changes under
4.	Start publishing them quarterly StatBank	YES	
5.	Document and update our web-page "About the statistics" with more detailed information about each of the four issues described as non-sampling errors	Not yet	This is an continuous ongoing process based on information mostly form other projects.
6.	Refine the "Quarterly accuracy report" to EUROSTAT with the new figures.	YES	See attachment
7.	Documentation and reporting to EUROSTAT	YES	

Item 3, the list of S.E. and CV tables are revised a little after a discussion with our Division for statistical methods. For estimates of change we only publish absolute and not relative standard errors. (This is in line documents from the Group of experts on LFS Precision requirements as fare as we can interpret them.)

S.E. and CV tables for county divided figures are also dropped in our quarterly release, but county divided CV for annual averages are included in chapter 4.

On the other side, we will publish gender and region-divided S.E. and CV figures of quarterly and annual levels.

About item 5 concerning updating of our web-page "About the statistics" with more detailed information about each of the four issues described as non-sampling errors. Another project about standardized non-response treatment in Statistics Norway are ongoing and will produce more certain information that may be important to dissemination of official information about these issues in our "About the statistics". So we will await new and better information about these issues before we update our "About the statistics". Bjørnstad (2013) documents sub-project 1, which is an overview of non-response and processing of non-response today in different surveys, NLFS among others. Chapter 5 in this report summaries some important findings.

References

Bjørnstad, Jan (editor) (2013): Standardisert frafallshåndtering for person- og husholdningsundersøkelser - Delprosjekt 1: Oversikt over status om frafall og frafallsbehandling, Interne dokumenter 6/2013, SSB (Norwegian only)

Hagesæther, Nina og Zhang, Li-Chun (2007) Om estimeringsusikkerhet og utvalgsplan i AKU Notat 2007/22, SSB. (Norwegian only)

Hagesæther, Nina and Zhang, Li-Chun (2009): A Note on the Effect of Auxiliary Information on the Variance of Cluster Sampling, JOS Vol. 25, No.3 2009.

Heldal, Johan (2000): Kalibrering av AKU. Dokumentasjon av metode og program. Notat 2000/7, SSB. (Norwegian only)

Thomsen, Ib and Villund, Ole (2011): Using Register Data to Evaluate the Effects of Proxy Interviews in the Norwegian Labour Force Survey JOS: March, Vol.27, No.1, 2011

Thomsen, Ib and Villund, Ole (2012): Nonresponse in the Norwegian labour force survey (LFS): Using administrative information to describe trends. Documents 54/2012, Statistics Norway

Vedø, Anne and Rafat, Dina (2003): Sammenligning av utvalgsplaner i AKU. Notater 56-2003, SSB.

Villund, Ole (2010): Household data in the Norwegian Labour Force Survey. Documents 2010/46, SSB.

Zhang, Li-Chun (1998): Dokumentasjonsrapport: Dokumentasjonsrapport: Den nye estimeringsmetoden for Arbeidskraftundersøkelsen (AKU): fylkesvis kalibrering med landsetterstratifiserte vekter som startverdier. Notater 1998/1, SSB. (norwegian only)

Zhang, Li-Chun (2000): Post-Stratification and Calibration - A Synthesis. The American Statistician, Vol. 54, No. 3 (Aug., 2000), pp. 178-184.

Zhang, Li-Chun (2005): On the Bias in Gross Labour Flow Estimates Due to Nonresponse and Misclassification, JOS Vol. 21, No.4 2005.

Zhang, Li-Chun (2006): A simplistic approach to variance estimation for calibrated estimators subjected to weighting adjustments. Unpublished document.

Zhang, Li-Chun (2008): On Some Common Practices of Systematic Sampling. Journal of Official Statistics, Vol.24, No.4, 2008. pp. 557–5 (www.jos.nu/Articles/abstract.asp?article=244557)

Li-Chun Zhang, Ib Thomsen and Øyvin Kleven (2013): On the Use of Auxiliary and Paradata for Dealing With Non-sampling Errors in Household Surveys. International Statistical Review (2013), 00, 0, 1–19 doi:10.1111/insr.12009

Steel, D. and McLaren, C. (2008): Design and Analysis of Repeated Surveys. Centre for Statistical and Survey Methodology, University of Wollongong, Working Paper 11-08.

List of figures

Figure 4.1. Coefficient of variation for employed persons by age 14
Figure 4.2. Coefficient of variation for unemployed persons by age and sex..... 14
Figure 4.3. Coefficient of variation for part-time employed persons aged 15-74, by sex 15
Figure 4.4. Coefficient of variationfor part-time employed persons, by age and sex 15
Figure 4.5. Standard error for the change from previous quarter in the number of employed persons, by age groups (in 1 000)..... 16
Figure 4.6 Standard error for the change from previous quarter in the number of unemployed persons, by age groups (in 1 000)..... 16
Figure 5.1. Yearly average of quarterly response rate in the Norwegian LFS 20

List of tables

Table 3.1. Relationship between sampling proportions in different counties 9
Table 4.1. Standard error for the estimated annual average of the proportion of unemployed to the population aged 15 to 74, by year and region 17
Table 4.2. Standard error for the difference in the estimated proportion of unemployment to the population aged 15 to 74 between two successive quarters 18
Table 4.3. Coefficient of variation (CV) for quarterly estimates. 2012..... 19
Table 4.4. Coefficient of variation (CV) for annual estimates, by region. 2012..... 19
Table 4.5. Coefficient of variation (CV) of annual estimates by county (NUTS3). 2012..... 19
Table 4.6. Table of action and fulfilment..... 23

B

Returadresse:
Statistisk sentralbyrå
NO-2225 Kongsvinger

Avsender:
Statistisk sentralbyrå

Postadresse:
Postboks 8131 Dep
NO-0033 Oslo

Besøksadresse:
Kongens gate 6, Oslo
Oterveien 23, Kongsvinger

E-post: ssb@ssb.no
Internett: www.ssb.no
Telefon: 62 88 50 00

ISBN 978-82-537-8791-6 (trykt)
ISBN 978-82-537-8792-3 (elektronisk)
ISSN 1891-5906

ISBN 978-82-537-8791-6

9 788253 787916

Statistisk sentralbyrå
Statistics Norway

Statistisk sentralbyrå
46/2013
Improved calculation and dissemination of CV in LF

Design: Siri Boquist