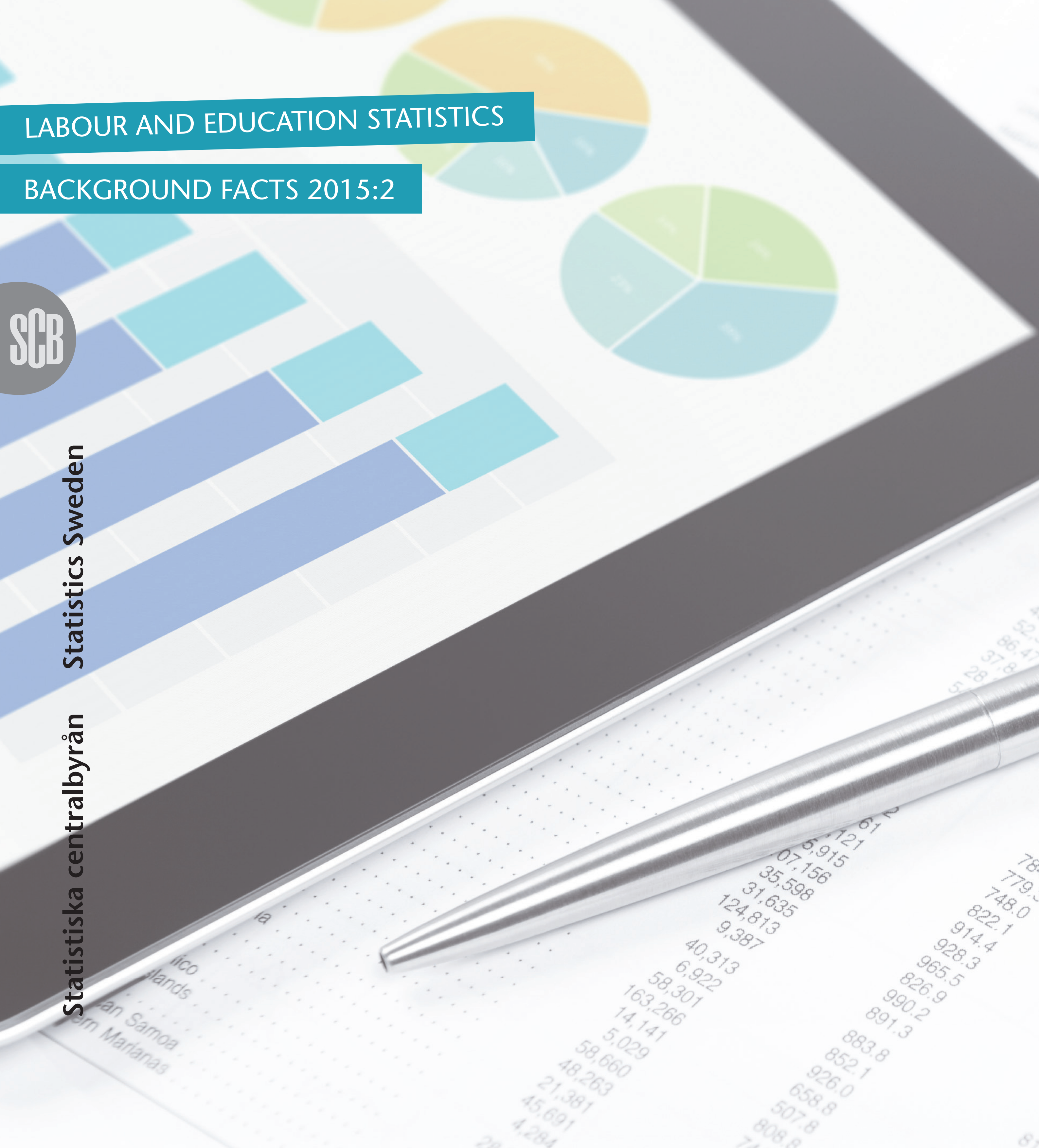

LABOUR AND EDUCATION STATISTICS

BACKGROUND FACTS 2015:2

Statistiska centralbyrån Statistics Sweden

Estimating household-related parameters in the Swedish LFS – theory and practice

I serien Bakgrundsfakta presenteras bakgrundsmaterial till den statistik som SCB producerar inom området arbetsmarknad och utbildning. Det kan röra sig om produktbeskrivningar, metodredovisningar samt olika sammanställningar som kan ge en överblick och underlätta användandet av statistiken.

Utgivna publikationer från 2000 i serien

Bakgrundsfakta till arbetsmarknads- och utbildningsstatistiken

2000:1 Övergång till yrkeskodning på fyrsiffernivå (SSYK) och införande av jobbstatus- kod i SCB:s lönestatistik

2000:2 The Information System for Occupational Injuries and the Work-related Health Problems Survey – A comparative study

2000:3 Konferens om utbildningsstatistik den 23 mars 2000

2001:1 Avvikelser i lönesummestatistiken – en jämförelse mellan LAPS och LSUM

2001:2 En longitudinell databas kring utbildning, inkomst och sysselsättning 1990–1998

2001:3 Staff training costs 1994–1999

2001:4 Studieresultat i högskolan i form av avklarade poäng

2001:5 Urvals- och estimationsförfarandet i de svenska arbetskraftsundersökningarna (AKU)

2001:6 Svar, bortfall och representativitet i Arbetsmiljöundersökningen 1999

2001:7 Individ- och företagsbaserad sysselsättningsstatistik – en jämförelse mellan AKU och KS

2002:1 Tidsseriebrott i utbildningsregistret 2001-01-01

2002:2 En longitudinell databas kring utbildning, inkomst och sysselsättning (LOUISE) 1990–1999

2003:1 Exempel på hur EU:s "Quality Reports" kan skrivas – avser Labour Cost Survey (LSC) 2000

2003:2 Förändrad redovisning av högskolans personal

2003:3 Individ- och företagsbaserad sysselsättningsstatistik – en fortsatt jämförelse mellan AKU och KS

2003:4 Sjukfrånvarande enligt SCB och sjukskrivna enligt RFV

2003:5 Informationssystemet om arbetsskador och undersökningen om arbetsorsakade besvär. En jämförande studie

2004:1 Samlad statistik från SCB avseende ohälsa

2004:2 Översyn av forskarutbildningsstatistiken. Bedömning av kvaliteten

2004.3 Sjukfrånvaro och ohälsa i Sverige – en belysning utifrån SCB:s statistik

2005:1 En longitudinell databas kring utbildning, inkomst och sysselsättning (LOUISE) 1990–2002

2005:2 Nordisk pendlingskarta. Huvudrapport

2005:3 Nordisk pendlingskarta. Delrapport 1–4.

2005:4 Flödesstatistik från AKU

2005:5 Flow statistics from the Swedish Labour Force Survey

2006:1 Sysselsättningsavgränsning i RAMS – Metodöversyn 2005

BACKGROUND FACTS

Estimating household-related parameters in the Swedish LFS – theory and practice

Labour and Education Statistics 2015:2

Statistics Sweden
2015

Background Facts
Labour and Education Statistics 2015:2

Estimating household-related parameters in the Swedish LFS – theory and practice

Statistics Sweden
2015

Previous publication – listed at the inside of the cover

Producer Statistics Sweden, Population and Welfare Department
SE-701 89 Örebro
+46 19 17 60 00

Enquiries Anna Broman +46 8 506 944 62
anna.broman@scb.se

Martin Axelson + 46 19 17 61 18
martin.axelson@scb.se

Cover: Ateljén, SCB. *Photo*: iStock

ISSN 1654-465X (Online)
ISSN 1103-7458 (Print)
ISBN 978-91-618-1628-6 (Print)
URN:NBN:SE:SCB-2015-AM76BR1502_pdf

Printed in Sweden
SCB-Tryck, Örebro 2015.08

Foreword

The objective of the Swedish labour force survey (LFS) is to describe the current employment conditions for the Swedish population and to give information on the development of the labour market at the individual level. Hence, the basis of the LFS is a sample of individuals in the working age. However, today the LFS is fully compliant with relevant EU-regulations, the most important being Council Regulation (EC) No 577/98 and Commission Regulation (EC) No 430/2005, as well as guidelines and recommendations issued by the International Labour Organization. Thus, part of the monthly LFS-sample, is used for identifying households, for which data are collected both on household-level characteristics, which are the same for all members of the household, and individual-level characteristics, most notably labour market conditions for individuals aged 15-74 years. Data on households are delivered to Eurostat annually.

Despite the availability of LFS-data at the household-level, Statistics Sweden has so far not used them for production of statistics. However, there are advanced plans for yearly publication of household-related LFS statistics from 2015 and onwards. This report presents, in a fairly detailed manner, methodology which may be used for deriving theoretically justified point and variance estimates, while at the same time fulfilling the consistency constraint imposed by Commission Regulation (EC) No 430/2005. To facilitate the implementation of the theoretical results, a set of SAS-macros have been derived, collectively named HUUVA 1.0.

The work was carried out by a project team consisting of Martin Axelson, who contributed extensively to the theoretical work and who is the author of this report, and Claes Andersson, who, in addition to being an important catalyst in the theoretical work, single-handedly derived the software HUUVA 1.0.

Despite being derived with the LFS in mind, the results presented are valid for any sample survey that matches the generic sampling design discussed in section 2. Moreover, the results presented can easily be extended to cover also point and variance estimation for estimators of change over time. Thus, the theoretical results presented should prove relevant and useful in a more general context as well.

Statistics Sweden June 2015

Inger Eklund

Hassan Mirza

A note of thanks

We would like to express appreciation to our survey respondents – the people, enterprises, government agencies and other institutions of Sweden – with whose cooperation Statistics Sweden is able to provide reliable and timely statistical information meeting the current needs of our modern society.

Contents

1 Introduction

The Swedish labour force survey (LFS), which has been conducted since the 1960-ies, is a sample survey targeting individuals within the working age. The objective of the LFS is to describe the current employment conditions for the population and to give information on the development of the labour market. Since 1970, the survey is conducted monthly, allowing for the production of monthly, quarterly and annual statistics with focus on both the number and the percentage of employed and unemployed persons. The LFS is the only source with continuous information on total unemployment and this information represents the official unemployment rate.

Today, the LFS is fully compliant with relevant EU-regulations, the most important being Council Regulation (EC) No 577/98 and Commission Regulation (EC) No 430/2005, as well as guidelines and recommendations issued by the International Labour Organization (2013). This means that in addition to collecting data referring to individuals aged 15-74 years, which is the main aim of the survey from a national perspective, part of the monthly LFS-sample is also used for identifying households for which data are collected. For households, data are collected both on household-level characteristics, which are the same for all members of the household, and individual-level characteristics, most notably labour market conditions for individuals aged 15-74 years. These may vary between members within a given household. Data on households, including weights to be used for estimation of population-level parameters, are delivered to Eurostat annually.

Despite the fact that Statistics Sweden for a number of years have collected household-level data and derived weights, the agency has so far refrained from using the material for producing and publishing national statistics. This is, however, about to change – there are advanced plans for yearly publication of household-related LFS statistics from 2015 and onwards. The purpose of this report is to describe, in a fairly detailed manner, the methodology used for calculating point and variance estimates of the type that will be produced. Most of the results presented are general, in the sense that they apply to any survey which shares the same design-features that underlie the LFS.

1.1 Outline of the report

As the LFS-design is rather complex, and additional complexity is added by existing EU-regulations, e.g., by constraints regarding numerical consistency between statistics based on household data and official LFS statistics, the content of the report is bound to be very technical in nature. However, rather than introducing the full problem from the beginning, the presentation follows a structure under which additional complexity is introduced in a stepwise manner. Section 2 provides a description of a generic sampling design which allows for sample rotation using so-called panels, thus introducing important notation which will be used throughout the report. Point estimation is discussed in section 3. An estimator under full response is presented in section 3.1, a non-response adjusted version of which is presented in section 3.2. In section 4, variance estimation under non-response is discussed, and an estimator for the variance of the point estimator presented in section 3.2 is derived. How to apply the estimators derived in sections 3 and 4 for estimation of household-related LFS statistics is discussed in section 5. An overview of the important features of the LFS sampling design is given in section 5.1. In section 5.2, estimation using HUUVA 1.0, a set of SAS-macros derived to facilitate the nontrivial issue of actually implementing the derived estimators for the LFS, is discussed. The report is concluded with some final remarks in section 6.

2 Sampling design for a rotating panel survey

The LFS is often presented as a panel survey with a rotating sample. In essence, this means that it is based on a sampling design which at the same time allows for (a) a sample of individuals to be followed over time, with data at the element level collected over time, and (b) the constitution of the sample to change over time in a controlled manner. In this section, a generic sampling design for implementing a rotating panel survey is introduced. Although not LFS-specific, the design includes all the important features of the LFS-design. A more detailed description of the LFS-design is deferred to section 5.1.

Let $U = \{1,...,k,...,N\}$ denote a population, the constitution of which is assumed to remain unchanged over the time-period in question. Let $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)},...,x_P^{(i)})'$ denote an auxiliary vector at time i, which takes on the value $\mathbf{x}_k^{(i)} = (x_{1k}^{(i)}, x_{2k}^{(i)},...,x_{Pk}^{(i)})'$ for element $k \in U$, $i = 1,...,I$. In what follows, for any set $S \subseteq U$ and any variable γ, $\sum_S \gamma_k$ is abbreviated notation for $\sum_{k \in S} \gamma_k$. Thus, the population total for $\mathbf{x}^{(i)}$ is given by

$$\mathbf{t}_{\mathbf{x}}^{(i)} = \sum_U \mathbf{x}_k^{(i)}$$

In what follows, $\mathbf{t}_{\mathbf{x}}^{(i)}$ is assumed known for $i = 1,...,I$. Let $\mathbf{y} = (y_1, y_2,..., y_Q)'$ and z denote two study variables, which for element $k \in U$ take on the values $\mathbf{y}_k^{(i)} = (y_{1k}^{(i)}, y_{2k}^{(i)},..., y_{Qk}^{(i)})'$ and $z_k^{(i)}$ at time i, $i = 1,...,I$. The population totals of $\mathbf{y}$ and z at time $i = 1,2,...,I$ are thus given by

$$\mathbf{t}_{\mathbf{y}}^{(i)} = \sum_U \mathbf{y}_k^{(i)} \tag{2.1}$$

and

$$t_z^{(i)} = \sum_U z_k^{(i)} \tag{2.2}$$

Although $\mathbf{x}$, $\mathbf{y}$ and z may vary at the element level over time, $\mathbf{x}_k^{(i)}$, $\mathbf{y}_k^{(i)}$ and $z_k^{(i)}$ represent fixed values at time i, $i = 1,...,I$. Let $\boldsymbol{\lambda}_q$ denote a $Q \times 1$ vector, which takes on the value 1 on row *q*, and 0 on all other rows. Then

$$t_{y_q}^{(i)} = \sum_U y_{qk}^{(i)} = \mathbf{t}_{\mathbf{y}}^{(i)\prime} \boldsymbol{\lambda}_q$$

At time i, $i=1,...,I$, the population U may be partitioned into $N_C^{(i)}$ clusters, denoted $U_1^{(i)},...,U_{N_C^{(i)}}^{(i)}$. Let $U_C^{(i)}=\{1,...,j,...,N_C^{(i)}\}$ denote the set of clusters and let $N_j^{(i)}$ denote the size of cluster j, $j\in U_C^{(i)}$, $i=1,...,I$. For example, a population of individuals may simultaneously be viewed as a population of households. As indicated by the notation, the partitioning of the population into clusters may vary over time even though the population, in terms of elements, does not. Let $j(k)$ denote the cluster to which element k belongs and let $c_k^{(i)}$, $k\in U$, be constants such that $\sum_{U_{j(k)}^{(i)}} c_l^{(i)}=1$ for each $j\in U_C^{(i)}$. A choice often encountered in practice is $c_k^{(i)}=1/N_{j(k)}^{(i)}$, but other possibilities exist. Now, let

$$\mathbf{Y}_k^{(i)}=c_k^{(i)}\sum\nolimits_{U_{j(k)}^{(i)}}\mathbf{y}_l^{(i)} \qquad (2.3)$$

and

$$Z_k^{(i)}=c_k^{(i)}\sum\nolimits_{U_{j(k)}^{(i)}}z_l^{(i)} \qquad (2.4)$$

It is a matter of algebra to show that alternative expressions for (2.1) and (2.2), which will prove useful for the purpose of this report, are given by

$$\mathbf{t}_{\mathbf{y}}^{(i)}=\mathbf{t}_{\mathbf{Y}}^{(i)}=\sum\nolimits_U\mathbf{Y}_k^{(i)} \qquad (2.5)$$

and

$$t_z^{(i)}=t_Z^{(i)}=\sum\nolimits_U Z_k^{(i)} \qquad (2.6)$$

Through appropriate definitions of $\mathbf{Y}_k^{(i)}$ or $Z_k^{(i)}$, $k\in U$, either (2.5) or (2.6) may readily be used to define parameters defined at the cluster-level. For example, $z_l^{(i)}=c_l^{(i)}$ for $l\in U_{j(k)}^{(i)}$ results in $Z_k^{(i)}=c_k^{(i)}$, which yields

$$t_Z^{(i)}=\sum\nolimits_U Z_k^{(i)}=\sum\nolimits_U c_k^{(i)}=\sum_{j\in U_C^{(i)}}\sum\nolimits_{U_j^{(i)}}c_k^{(i)}=\sum_{j\in U_C^{(i)}}1=N_C^{(i)}$$

i.e., the number of clusters at time i.

Let $a^{(i)}$, $i=1,...,I$, denote pre-defined constants. The parameters of interest are

$$T_{y_q}=\sum_{i=1}^{I}a^{(i)}t_{y_q}^{(i)}=\mathbf{T}_{\mathbf{y}}'\boldsymbol{\lambda}_q \qquad (2.7)$$

$q = 1,...,Q$, where

$$\mathbf{T}_{\mathbf{y}} = \sum_{i=1}^{I} a^{(i)} \mathbf{t}_{\mathbf{y}}^{(i)}$$

and

$$T_z = \sum_{i=1}^{I} a^{(i)} t_z^{(i)} \qquad (2.8)$$

In order to estimate (2.7) and (2.8), a total of $V^* = I + V - 1$ independent subsamples from U will be used. Let s_v denote the v:th subsample, of size such n_v, drawn from U according to the sampling design $p_v(\cdot)$, with corresponding first- and second-order inclusion probabilities π_{vk} and π_{vkl}, $v = 1,...,V^*$. It is assumed that the designs and the sample sizes are such that $\pi_{vkl} > 0$ for all $\{k,l\} \in U$, $v = 1,...,V^*$. At each time, exactly V subsamples will be subject to data collection. More specifically, at time i, $i = 1,...,I$, the intention is to collect the following information for estimation of the parameters of interest:

- $\mathbf{x}_k^{(i)}$ and $\mathbf{y}_k^{(i)}$ for $k \in s_v$, $v = i,...,i + V - 1$
- $\mathbf{Y}_k^{(i)} = c_k^{(i)} \sum_{U_{j(k)}^{(i)}} \mathbf{y}_l^{(i)}$ and $Z_k^{(i)} = c_k^{(i)} \sum_{U_{j(k)}^{(i)}} z_l^{(i)}$ for $k \in s_v$, $v = i$

At time i, the elements $k \in s_v$, $v = i,...,i + V - 1$ are said to be directly sampled, whereas the elements $l \in U_{j(k)}^{(i)}$, $l \neq k$, for $k \in s_v$, $v = i$, are said to be indirectly sampled, $i = 1,...,I$. It should be noted that in the notation introduced above, the sample numbers for which data are to be collected at time i are expressed as functions of time i. In Appendix 1, an example of the data to be collected for the situation where $I = 3$ and $V = 3$ is given.

3 Point estimation

3.1 Estimation under full response

Suppose the parameters of interest are T_{y_q} and T_z, as defined by (2.7) and (2.8), and that $V^* = I + V - 1$ independent subsamples from U are drawn according to the generic design presented in section 2. Theoretically, an estimator for $\mathbf{T_Y}$ under full response could be based on

- $\mathbf{x}_k^{(i)}$ for $k \in s_v$, $v = i,...,i+V-1$
- $\mathbf{y}_k^{(i)}$ for $k \in s_v$, $v = i+1,...,i+V-1$
- $\mathbf{y}_l^{(i)}$ for $l \in U_{j(k)}^{(i)}$ and $k \in s_v$, $v = i$

$i = 1,...,I$, i.e., on information collected from both directly and indirectly sampled elements. However, in what follows, it is assumed that only data on directly sampled elements will be used. The reason for considering this less than ideal scenario, is that the parameter $\mathbf{t}_\mathbf{y}^{(i)}$ is assumed to be estimated at time $i = 1,...,I$, at a point in time when only data on directly sampled elements are available. This scenario closely resembles the situation encountered in practice in the LFS, where monthly estimates are based solely on data from directly sampled individuals. Moreover, although theoretically possible, it is far from straightforward to construct an estimator $\mathbf{t}_\mathbf{y}^{(i)}$, $i = 1,...,I$, which under the design in question (a) makes combined use of data from both directly and indirectly sampled elements and (b) is relatively easy to implement.

For estimation of T_z, which will be estimated at time I, it is assumed that under full response, the following data would be available for time $i = 1,...,I$ at the time of estimation:

- $\mathbf{x}_k^{(i)}$ and $\mathbf{y}_k^{(i)}$ for $k \in s_v$, $v = i,...,i+V-1$
- $\mathbf{Y}_k^{(i)} = c_k^{(i)} \sum_{U_{j(k)}^{(i)}} \mathbf{y}_l^{(i)}$ and $Z_k^{(i)} = c_k^{(i)} \sum_{U_{j(k)}^{(i)}} z_l^{(i)}$ for $k \in s_v$, $v = i$

For $i = 1,...,I$ and $v = 1,...,V^*$, let $b_{vk}^{(i)}$ be predefined constants such that $\sum_{v=i}^{i+V-1} b_{vk}^{(i)} = 1$ and $b_{vk}^{(i)} = 0$ for $v < i$ and $i+V-1 < v$. One possible choice for $b_{vk}^{(i)}$, $i = 1,...,I$, is

$$b_{vk}^{(i)} = \begin{cases} \pi_{vk} / \sum_{v'=i}^{i+V-1} \pi_{v'k} & v = i,...,i+V-1 \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

If the designs $p_v(\cdot)$, $v = 1,...,V^*$, are such that $\pi_{vk} = \pi_k$ for each $k \in U$, (3.1) simplifies to

$$b_{vk}^{(i)} = b_v = \begin{cases} 1/V & v = i,...,i+V-1 \\ 0 & \text{otherwise} \end{cases}$$

Let

$$\hat{\mathbf{t}}_{ayc,s}^{(i)} = [(\mathbf{t}_{\mathbf{x}}^{(i)} - \hat{\mathbf{t}}_{a\mathbf{x},s}^{(i)})'\hat{\mathbf{B}}_{a\mathbf{y},s}^{(i)}]' + \hat{\mathbf{t}}_{a\mathbf{y},s}^{(i)}$$

where

$$\hat{\mathbf{t}}_{a\mathbf{x},s}^{(i)} = \sum_{v=i}^{i+V-1} \sum\nolimits_{s_v} \frac{b_{vk}^{(i)} \mathbf{x}_k^{(i)}}{\pi_{vk}}$$

$$\hat{\mathbf{t}}_{a\mathbf{y},s}^{(i)} = \sum_{v=i}^{i+V-1} \sum\nolimits_{s_v} \frac{b_{vk}^{(i)} \mathbf{y}_k^{(i)}}{\pi_{vk}}$$

and

$$\hat{\mathbf{B}}_{a\mathbf{y},s}^{(i)} = (\sum_{v=i}^{i+V-1} \sum\nolimits_{s_v} \frac{b_{vk}^{(i)} \mathbf{x}_k^{(i)} \mathbf{x}_k^{(i)'}}{\pi_{vk}})^{-1} \sum_{v=i}^{i+V-1} \sum\nolimits_{s_v} \frac{b_{vk}^{(i)} \mathbf{x}_k^{(i)} \mathbf{y}_k^{(i)'}}{\pi_{vk}}$$

In Statistics Sweden (2014a, p. 27), it is shown that $\hat{\mathbf{t}}_{ayc,s}^{(i)}$ is approximately unbiased for $\mathbf{t}_{\mathbf{y}}^{(i)}$. $i = 1,...,I$, given that n_v, $v = 1,...,V^*$, is large enough. Thus, under full response, an approximately unbiased estimator for T_{y_q}, $q = 1,...,Q$, is given by

$$\hat{T}_{y_q,s} = \hat{\mathbf{T}}_{\mathbf{y},s}{}' \boldsymbol{\lambda}_q \tag{3.2}$$

where

$$\hat{\mathbf{T}}_{\mathbf{y},s} = \sum_{i=1}^{I} a^{(i)} \hat{\mathbf{t}}_{ayc,s}^{(i)}$$

For T_z defined by (2.8), an estimator under full response is given by

$$\hat{T}_{z,s} = \sum_{i=1}^{I} a^{(i)} \hat{t}_{zc,s}^{(i)} \tag{3.3}$$

where

$$\begin{aligned} \hat{t}_{zc,s}^{(i)} &= (\hat{\mathbf{t}}_{ayc,s}^{(i)} - \hat{\mathbf{t}}_{\mathbf{Y},s}^{(i)})'\hat{\mathbf{B}}_{z,s}^{(i)} + \hat{t}_{Z,s}^{(i)} \\ &= (\mathbf{t}_{\mathbf{x}}^{(i)} - \hat{\mathbf{t}}_{a\mathbf{x},s}^{(i)})'\hat{\mathbf{B}}_{az,s}^{(i)} + (\hat{\mathbf{t}}_{a\mathbf{y},s}^{(i)} - \hat{\mathbf{t}}_{\mathbf{Y},s}^{(i)})'\hat{\mathbf{B}}_{z,s}^{(i)} + \hat{t}_{Z,s}^{(i)} \end{aligned} \tag{3.4}$$

with

$$\hat{\mathbf{t}}^{(i)}_{\mathbf{Y},s} = \sum_{v=i}^{i} \sum_{s_v} \frac{\mathbf{Y}^{(i)}_k}{\pi_{vk}}$$

$$\hat{\mathbf{B}}^{(i)}_{z,s} = (\sum_{v=i}^{i} \sum_{s_v} \frac{\mathbf{Y}^{(i)}_k \mathbf{Y}^{(i)'}_k}{\pi_{vk}})^{-1} \sum_{v=i}^{i} \sum_{s_v} \frac{\mathbf{Y}^{(i)}_k Z^{(i)}_k}{\pi_{vk}}$$

$$\hat{t}^{(i)}_{Z,s} = \sum_{v=i}^{i} \sum_{s_v} \frac{Z^{(i)}_k}{\pi_{vk}}$$

and $\hat{\mathbf{B}}^{(i)}_{az,s} = \hat{\mathbf{B}}^{(i)}_{ay,s} \hat{\mathbf{B}}^{(i)}_{z,s}$.

Given that the sample size n_v, $v = 1,...,V^*$, is large, it follows that

$$\hat{t}^{(i)}_{zc,s} \approx (\mathbf{t}^{(i)}_{\mathbf{x}} - \hat{\mathbf{t}}^{(i)}_{\mathbf{ax},s})' E(\hat{\mathbf{B}}^{(i)}_{az,s}) + (\hat{\mathbf{t}}^{(i)}_{ay,s} - \hat{\mathbf{t}}^{(i)}_{\mathbf{Y},s})' E(\hat{\mathbf{B}}^{(i)}_{z,s}) + \hat{t}^{(i)}_{Z,s}$$

Since $E(\hat{\mathbf{t}}^{(i)}_{\mathbf{ax},s}) = \mathbf{t}^{(i)}_{\mathbf{x}}$, $E(\hat{\mathbf{t}}^{(i)}_{ay,s}) = E(\hat{\mathbf{t}}^{(i)}_{\mathbf{Y},s})$, and $E(\hat{t}^{(i)}_{Z,s}) = t^{(i)}_z$, it follows that $\hat{t}^{(i)}_{zc,s}$ is approximately unbiased for $t^{(i)}_z$ when the sample size is large enough. This also implies that $\hat{T}_{z,s}$ is approximately unbiased for T_z under the same conditions.

Suppose $z^{(i)}_k$ is in the row-space of $\mathbf{y}^{(i)}_k$, i.e., that there exists a constant vector $\boldsymbol{\lambda}$ such that $z^{(i)}_k = \mathbf{y}^{(i)'}_k \boldsymbol{\lambda}$ for all $k \in U$ and $i = 1,..., I$. As this implies that $Z^{(i)}_k = \mathbf{Y}^{(i)'}_k \boldsymbol{\lambda}$, it follows that $\hat{\mathbf{B}}^{(i)}_{z,s} = \boldsymbol{\lambda}$ and $\hat{t}^{(i)}_{Z,s} = \hat{\mathbf{t}}^{(i)'}_{\mathbf{Y},s} \boldsymbol{\lambda}$. Hence, (3.4) simplifies to

$$\hat{t}^{(i)}_{zc,s} = (\hat{\mathbf{t}}^{(i)}_{ayc,s} - \hat{\mathbf{t}}^{(i)}_{\mathbf{Y},s})' \hat{\mathbf{B}}^{(i)}_{z,s} + \hat{t}^{(i)}_{Z,s} = (\hat{\mathbf{t}}^{(i)}_{ayc,s} - \hat{\mathbf{t}}^{(i)}_{\mathbf{Y},s})' \boldsymbol{\lambda} + \hat{\mathbf{t}}^{(i)'}_{\mathbf{Y},s} \boldsymbol{\lambda} = \hat{\mathbf{t}}^{(i)'}_{ayc,s} \boldsymbol{\lambda}$$

which in turn means that (3.3) simplifies to

$$\hat{T}_{z,s} = \sum_{i=1}^{I} a^{(i)} \hat{\mathbf{t}}^{(i)'}_{ayc,s} \boldsymbol{\lambda} = \hat{\mathbf{T}}'_{\mathbf{y},s} \boldsymbol{\lambda}$$

Thus, for $z^{(i)}_k = y^{(i)}_{qk} = \mathbf{y}^{(i)'}_k \boldsymbol{\lambda}_q$, $q = 1,..., Q$, (3.2) and (3.3) will produce identical point estimates.

3.2 Estimation under nonresponse

Due to nonresponse, neither (3.2) nor (3.3) can be used in practice. In this section, nonresponse adjusted versions are derived. Let $r_v^{(i)}$ denote the subset of units in s_v which are treated as responding units at time i, i.e., $r_v^{(i)} \subseteq s_v$, $i = 1,...,I$ and $v = 1,...,V^*$. For $k \in s_v$, $v = 1,...,V^*$, let $R_{vk}^{(i)}$, $i = 1,...,I$, be defined as

$$R_{vk}^{(i)} = \begin{cases} 0 & \text{if } k \notin r_v^{(i)}, i \le v \le i + V - 1 \\ 1 & \text{otherwise} \end{cases}$$

In what follows, it is presumed that the response behaviour is governed by a stochastic response distribution, RD, such that:

At time i, $i = 1,...,I$, element $k \in s_v$, $v = 1,...,V^*$, responds with probability

$$\Pr(k \in r_v^{(i)} \mid r_v^{(i-1)},...,r_v^{(1)}, s_v) = \Pr(R_{vk}^{(i)} = 1 \mid r_v^{(i-1)},...,r_v^{(1)}, s_v) = \theta_{vk}^{(i)}$$

Moreover, at time i, $i = 1,...,I$, elements $\{k,l\} \in s_v$, $v = 1,...,V^*$, $k \ne l$, respond independently of each other, i.e.,

$$\Pr(\{k,l\} \in r_v^{(i)} \mid r_v^{(i-1)},...,r_v^{(1)}, s_v) = \Pr(R_{vk}^{(i)} R_{vl}^{(i)} = 1 \mid r_v^{(i-1)},...,r_v^{(1)}, s_v) = \theta_{vk}^{(i)} \theta_{vl}^{(i)}$$

Any pair of elements k and l such that $k \in s_v$ and $l \in s_{v'}$, $v \ne v'$, respond independently at all times.

As indicated, the response distribution at time i, $i = 1,...,I$, may depend on the realized response sets at previous times, but for the sake of notational simplicity, the conditional arguments is dropped from θ. It will prove convenient to use $r_v^{(i)} = s_v$ and $\theta_{vk}^{(i)} = 1$ for $v < i$ and $i + V - 1 < v$, $i = 1,...,I$, i.e., if the set s_v is not subject to data collection at time i, all $k \in s_v$ will be treated as responding elements.

Assuming fully known response probabilities, a nonresponse adjusted version of (3.2) would be given by

$$\hat{T}_{y_q,h} = \sum_{i=1}^{I} a^{(i)} \hat{\mathbf{t}}_{ayc,h}^{(i)\prime} \boldsymbol{\lambda}_q$$

where

$$\hat{\mathbf{t}}_{ayc,h}^{(i)} = [(\mathbf{t}_{\mathbf{x}}^{(i)} - \hat{\mathbf{t}}_{a\mathbf{x},h}^{(i)})' \hat{\mathbf{B}}_{ay,h}^{(i)}]' + \hat{\mathbf{t}}_{ay,h}^{(i)}$$

with

$$\hat{\mathbf{t}}^{(i)}_{a\mathbf{x},h} = \sum_{v=i}^{i+V-1} \sum\nolimits_{r_v^{(i)}} \frac{b^{(i)}_{vk} \mathbf{x}^{(i)}_k}{\pi_{vk} \theta^{(i)}_{vk}} \tag{3.5}$$

$$\hat{\mathbf{t}}^{(i)}_{a\mathbf{y},h} = \sum_{v=i}^{i+V-1} \sum\nolimits_{r_v^{(i)}} \frac{b^{(i)}_{vk} \mathbf{y}^{(i)}_k}{\pi_{vk} \theta^{(i)}_{vk}} \tag{3.6}$$

and

$$\hat{\mathbf{B}}^{(i)}_{a\mathbf{y},h} = \Big(\sum_{v=i}^{i+V-1} \sum\nolimits_{r_v^{(i)}} \frac{b^{(i)}_{vk} \mathbf{x}^{(i)}_k \mathbf{x}^{(i)\prime}_k}{\pi_{vk} \theta^{(i)}_{vk}} \Big)^{-1} \sum_{v=i}^{i+V-1} \sum\nolimits_{r_v^{(i)}} \frac{b^{(i)}_{vk} \mathbf{x}^{(i)}_k \mathbf{y}^{(i)\prime}_k}{\pi_{vk} \theta^{(i)}_{vk}} \tag{3.7}$$

For (3.3), a nonresponse adjusted version would be given by

$$\hat{T}_{z,h} = \sum_{i=1}^{I} a^{(i)} \hat{t}^{(i)}_{zc,h}$$

where

$$\begin{aligned} \hat{t}^{(i)}_{zc,h} &= (\hat{\mathbf{t}}^{(i)}_{ayc,h} - \hat{\mathbf{t}}^{(i)}_{\mathbf{Y},h})' \hat{\mathbf{B}}^{(i)}_{z,h} + \hat{t}^{(i)}_{Z,h} \\ &= (\mathbf{t}^{(i)}_{\mathbf{x}} - \hat{\mathbf{t}}^{(i)}_{a\mathbf{x},h})' \hat{\mathbf{B}}^{(i)}_{az,h} + (\hat{\mathbf{t}}^{(i)}_{a\mathbf{y},h} - \hat{\mathbf{t}}^{(i)}_{\mathbf{Y},h})' \hat{\mathbf{B}}^{(i)}_{z,h} + \hat{t}^{(i)}_{Z,h} \end{aligned}$$

with

$$\hat{\mathbf{t}}^{(i)}_{\mathbf{Y},h} = \sum_{v=i}^{i} \sum\nolimits_{r_v^{(i)}} \frac{\mathbf{Y}^{(i)}_k}{\pi_{vk} \theta^{(i)}_{vk}} \tag{3.8}$$

$$\hat{\mathbf{B}}^{(i)}_{z,h} = \Big(\sum_{v=i}^{i} \sum\nolimits_{r_v^{(i)}} \frac{\mathbf{Y}^{(i)}_k \mathbf{Y}^{(i)\prime}_k}{\pi_{vk} \theta^{(i)}_{vk}} \Big)^{-1} \sum_{v=i}^{i} \sum\nolimits_{r_v^{(i)}} \frac{\mathbf{Y}^{(i)}_k Z^{(i)}_k}{\pi_{vk} \theta^{(i)}_{vk}} \tag{3.9}$$

$$\hat{t}^{(i)}_{Z,h} = \sum_{v=i}^{i} \sum\nolimits_{r_v^{(i)}} \frac{Z^{(i)}_k}{\pi_{vk} \theta^{(i)}_{vk}} \tag{3.10}$$

and $\hat{\mathbf{B}}^{(i)}_{az,h} = \hat{\mathbf{B}}^{(i)}_{a\mathbf{y},h} \hat{\mathbf{B}}^{(i)}_{z,h}$.

In practice, however, the response probabilities are unknown and have to be estimated. A commonly used approach is to make assumptions on the response distribution, in terms of an explicitly stated response model, and to derive estimator expressions under the assumption that the response model is identical to the true, but unknown, response distribution. In what follows, the following response model, based on response homogeneity groups, will be used:

- At time i, $i=1,...,I$, there exists a known partitioning $s_{vg}^{(i)}$, $g=1,...,G_v^{(i)}$, of s_v, $v=1,...,V^*$, such that

$$\theta_{vk}^{(i)}=\theta_{vg}^{(i)} \text{ for } k\in s_{vg}^{(i)},\; g=1,...,G_v^{(i)}$$

  The number of elements in $s_{vg}^{(i)}$ is denoted $n_{vg}^{(i)}$.

In order to simplify the notation, no conditional arguments will be used for the RHG-probabilities. For $i=1,...,I$ and $v=1,...,V^*$, let $\mathbf{m}_v^{(i)}$ denote the response count vector, i.e., $\mathbf{m}_v^{(i)}=(m_{v1}^{(i)},...,m_{vg}^{(i)},...,m_{vG_v^{(i)}}^{(i)})'$, where $m_{vg}^{(i)}$ denotes the number of elements in $r_{vg}^{(i)}=r_v^{(i)}\cap s_{vg}^{(i)}$. Since $r_v^{(i)}=s_v$ by definition for $v<i$ and $i+V-1<v$, it will prove convenient to use $G_v^{(i)}=1$, which yields $\mathbf{m}_v^{(i)}=n_v$, for $v<i$ and $i+V-1<v$.

The results below are derived, and justified, under the following two assumptions:

A1: The used response model coincides with the true response distribution.

A2: For $i=1,...,I$ and $v=1,...,V^*$, $m_{vg}^{(i)}\geq 2$ for $g=1,...,G_v^{(i)}$.

Under the response model, $\theta_{vk}^{(i)}=\theta_{vg}^{(i)}$ for all $k\in s_{vg}^{(i)}$. Hence, an unbiased estimator for $\theta_{vk}^{(i)}$ is given by $\hat{\theta}_{vk}^{(i)}=\hat{\theta}_{vg}^{(i)}=m_{vg}^{(i)}/n_{vg}^{(i)}$, $g=1,...,G_v^{(i)}$, $i=1,...,I$ and $v=1,...,V^*$. Using estimated response probabilities as plug-ins for the true, but unknown, response probabilities in (3.5) – (3.7), a working estimator for T_{y_q} is given by

$$\hat{T}_{y_q}=\sum_{i=1}^{I}a^{(i)}\hat{\mathbf{t}}_{ayc}^{(i)\prime}\boldsymbol{\lambda}_q \tag{3.11}$$

where

$$\hat{\mathbf{t}}_{ayc}^{(i)}=[(\mathbf{t}_{\mathbf{x}}^{(i)}-\hat{\mathbf{t}}_{a\mathbf{x}}^{(i)})'\hat{\mathbf{B}}_{ay}^{(i)}]'+\hat{\mathbf{t}}_{ay}^{(i)} \tag{3.12}$$

with

$$\hat{\mathbf{t}}_{a\mathbf{x}}^{(i)}=\sum_{v=i}^{i+V-1}\sum\nolimits_{r_v^{(i)}}\frac{b_{vk}^{(i)}\mathbf{x}_k^{(i)}}{\pi_{vk}\hat{\theta}_{vk}^{(i)}}$$

$$\hat{\mathbf{t}}_{a\mathbf{y}}^{(i)} = \sum_{v=i}^{i+V-1} \sum\nolimits_{r_v^{(i)}} \frac{b_{vk}^{(i)} \mathbf{y}_k^{(i)}}{\pi_{vk} \hat{\theta}_{vk}^{(i)}}$$

and

$$\hat{\mathbf{B}}_{a\mathbf{y}}^{(i)} = \left(\sum_{v=i}^{i+V-1} \sum\nolimits_{r_v^{(i)}} \frac{b_{vk}^{(i)} \mathbf{x}_k^{(i)} \mathbf{x}_k^{(i)\prime}}{\pi_{vk} \hat{\theta}_{vk}^{(i)}}\right)^{-1} \sum_{v=i}^{i+V-1} \sum\nolimits_{r_v^{(i)}} \frac{b_{vk}^{(i)} \mathbf{x}_k^{(i)} \mathbf{y}_k^{(i)\prime}}{\pi_{vk} \hat{\theta}_{vk}^{(i)}}$$

Using the same approach, using estimated response probabilities as plug-ins for the true, but unknown, response probabilities in (3.5) – (3.7) and (3.9) – (3.10), , a working estimator for T_z is given by

$$\hat{T}_z = \sum_{i=1}^{I} a^{(i)} \hat{t}_{zc}^{(i)} \tag{3.13}$$

where

$$\begin{aligned} \hat{t}_{zc}^{(i)} &= (\hat{\mathbf{t}}_{ayc}^{(i)} - \hat{\mathbf{t}}_{\mathbf{Y}}^{(i)})' \hat{\mathbf{B}}_z^{(i)} + \hat{t}_Z^{(i)} \\ &= (\mathbf{t}_{\mathbf{x}}^{(i)} - \hat{\mathbf{t}}_{a\mathbf{x}}^{(i)})' \hat{\mathbf{B}}_{az}^{(i)} + (\hat{\mathbf{t}}_{a\mathbf{y}}^{(i)} - \hat{\mathbf{t}}_{\mathbf{Y}}^{(i)})' \hat{\mathbf{B}}_z^{(i)} + \hat{t}_Z^{(i)} \end{aligned} \tag{3.14}$$

with

$$\hat{\mathbf{t}}_{\mathbf{Y}}^{(i)} = \sum_{v=i}^{i} \sum\nolimits_{r_v^{(i)}} \frac{\mathbf{Y}_k^{(i)}}{\pi_{vk} \hat{\theta}_{vk}^{(i)}}$$

$$\hat{\mathbf{B}}_z^{(i)} = \left(\sum_{v=i}^{i} \sum\nolimits_{r_v^{(i)}} \frac{\mathbf{Y}_k^{(i)} \mathbf{Y}_k^{(i)\prime}}{\pi_{vk} \hat{\theta}_{vk}^{(i)}}\right)^{-1} \sum_{v=i}^{i} \sum\nolimits_{r_v^{(i)}} \frac{\mathbf{Y}_k^{(i)} Z_k^{(i)}}{\pi_{vk} \hat{\theta}_{vk}^{(i)}}$$

$$\hat{t}_Z^{(i)} = \sum_{v=i}^{i} \sum\nolimits_{r_v^{(i)}} \frac{Z_k^{(i)}}{\pi_{vk} \hat{\theta}_{vk}^{(i)}}$$

and $\hat{\mathbf{B}}_{az}^{(i)} = \hat{\mathbf{B}}_{ay}^{(i)} \hat{\mathbf{B}}_z^{(i)}$. It is a matter of algebra to show that for $z_k^{(i)}$ is in the row-space of $\mathbf{y}_k^{(i)}$, i.e., for $z_k^{(i)} = \mathbf{y}_k^{(i)\prime} \boldsymbol{\lambda}$ for all $k \in U$ and $i = 1,\ldots,I$, where $\boldsymbol{\lambda}$ is a constant vector, (3.13) simplifies to

$$\hat{T}_z = \sum_{i=1}^{I} a^{(i)} \hat{\mathbf{t}}_{ayc}^{(i)\prime} \boldsymbol{\lambda} = \hat{\mathbf{T}}_{\mathbf{y}}' \boldsymbol{\lambda}$$

Hence, for $z_k^{(i)} = y_{qk}^{(i)} = \mathbf{y}_k^{(i)\prime}\boldsymbol{\lambda}_q$, $q = 1,...,Q$, (3.11) and (3.13) will produce identical point estimates. Considering the LFS, this is a very important feature, as it means that (3.13) can be used to fulfill the following consistency constraint, imposed by Commission Regulation (EC) No 430/2005:

> *Consistency between annual sub-sample totals and full sample annual averages shall be ensured for employment, unemployment and inactive population by sex and for the following age groups: 15 to 24, 25 to 34, 35 to 44, 45 to 54, 55 +.*

Under assumptions *A1* and *A2*, the set $r_{vg}^{(i)}$ is an SI- sample of size $m_{vg}^{(i)}$ from the $n_{vg}^{(i)}$ elements in $s_{vg}^{(i)}$, $g = 1,...,G_v^{(i)}$, $i = 1,...,I$ and $v = 1,...,V^*$, conditional on s_v, $\mathbf{m}_v^{(1)}$,…, $\mathbf{m}_v^{(i)}$. Thus, it follows that for $k \in s_{vg}^{(i)}$

$$E_{RD}(R_{vk}^{(i)} \mid \mathbf{m}_v^{(i)},...,\mathbf{m}_v^{(1)}, s_v) = \hat{\theta}_{vk}^{(i)} = \hat{\theta}_{vg}^{(i)} = m_{vg}^{(i)} / n_{vg}^{(i)}$$

which means that

$$\begin{aligned} E(\hat{\mathbf{t}}_{a\mathbf{x}}^{(i)}) &= \sum_{v=i}^{i+V-1} E_{p_v} E_{\mathbf{m}_v^{(1)}} .... E_{\mathbf{m}_v^{(i)}} E_{RD}(\sum\nolimits_{s_v} \frac{R_{vk}^{(i)} b_{vk}^{(i)} \mathbf{x}_k^{(i)}}{\pi_{vk}\hat{\theta}_{vk}^{(i)}} \mid \mathbf{m}_v^{(i)},...,\mathbf{m}_v^{(1)}, s_v) \\ &= \sum_{v=i}^{i+V-1} E_{p_v}(\sum\nolimits_{s_v} \frac{b_{vk}^{(i)} \mathbf{x}_k^{(i)}}{\pi_{vk}}) \\ &= \sum_{v=i}^{i+V-1} \sum\nolimits_U b_{vk}^{(i)} \mathbf{x}_k^{(i)} \\ &= \sum\nolimits_U \mathbf{x}_k^{(i)} \sum_{v=i}^{i+V-1} b_{vk}^{(i)} \\ &= \mathbf{t}_{\mathbf{x}}^{(i)} \end{aligned}$$

for $i = 1,...,I$. Analogously, it follows that $E(\hat{\mathbf{t}}_{a\mathbf{y}}^{(i)}) = E(\hat{\mathbf{t}}_{\mathbf{Y}}^{(i)}) = \mathbf{t}_{\mathbf{y}}^{(i)}$ and $E(\hat{t}_Z^{(i)}) = t_z^{(i)}$. Hence, given that the sizes of all samples and response sets are large enough for the approximation

$$\hat{t}_{zc}^{(i)} \approx (\mathbf{t}_{\mathbf{x}}^{(i)} - \hat{\mathbf{t}}_{a\mathbf{x}}^{(i)})'\mathbf{B}_{az}^{(i)} + (\hat{\mathbf{t}}_{a\mathbf{y}}^{(i)} - \hat{\mathbf{t}}_{\mathbf{Y}}^{(i)})'\mathbf{B}_z^{(i)} + \hat{t}_Z^{(i)}$$

where $\mathbf{B}_{az}^{(i)} = E(\hat{\mathbf{B}}_{az}^{(i)})$ and $\mathbf{B}_z^{(i)} = E(\hat{\mathbf{B}}_z^{(i)})$, to be valid, it follows that $E(\hat{t}_{zc}^{(i)}) \approx t_z^{(i)}$, and consequently, $E(\hat{T}_z) \approx T_z$. As $\hat{T}_z$ simplifies to $\hat{\mathbf{T}}_{\mathbf{y}}'\boldsymbol{\lambda}$ for $z_k^{(i)} = \mathbf{y}_k^{(i)\prime}\boldsymbol{\lambda}$, approximate unbiasedness of (3.13) implies that (3.11) is approximately unbiased for T_{y_q}, $q = 1,...,Q$.

Clearly, both (3.11) and (3.13) can be viewed as generalized regression estimators under two-phase sampling, with the second-phase design corresponding to stratified simple random sampling with strata defined by the response homogeneity groups. Thus, given that π_{vk} and $\hat{\theta}_{vk}^{(i)}$ can be obtained for $k \in r_v^{(i)}$, $i = 1,...,I$ and $v = 1,...,V^*$, the software ETOS 2.0 (Andersson, 2012), derived at Statistics Sweden, may be used to compute estimates of T_{y_q}, $q = 1,...,Q$, and T_z according to (3.11) and (3.13), respectively.

4 Variance estimation under nonresponse

4.1 Deriving an estimator

In this section, a working estimator for $V(\hat{T}_z)$, the variance of $\hat{T}_z$, is derived. As $\hat{T}_z$ simplifies to $\hat{T}_{y_q}$ when $z_k^{(i)} = y_{qk}^{(i)}$ for $i = 1,...,I$ and $q = 1,...,Q$, the same estimator can be used to derive an estimator for $V(\hat{T}_{y_q})$, $q = 1,...,Q$. For $i = 1,...,I$, let $d_{vk}^{(i)}$ and $e_{vk}^{(i)}$ be defined for every $k \in s_v$, $v = 1,...,V^*$, as

$$d_{vk}^{(i)} = \begin{cases} \mathbf{y}_k^{(i)\prime}\hat{\mathbf{B}}_z^{(i)} - \mathbf{x}_k^{(i)\prime}\hat{\mathbf{B}}_{az}^{(i)} & \text{if } R_{vk}^{(i)} = 1 \text{ and } i \le v \le i + V - 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$e_{vk}^{(i)} = \begin{cases} Z_k^{(i)} - \mathbf{Y}_k^{(i)\prime}\hat{\mathbf{B}}_z^{(i)} & \text{if } R_{vk}^{(i)} = 1 \text{ and } v = i \\ 0 & \text{otherwise} \end{cases}$$

with $\hat{\mathbf{B}}_z^{(i)}$ and $\hat{\mathbf{B}}_{az}^{(i)}$ according to the definitions in section 3.2. In Appendix 2, an example of $d_{vk}^{(i)}$ and $e_{vk}^{(i)}$ for the situation where $I = 3$ and $V = 3$ is given.

It is a matter of algebra to show that $\hat{t}_{zc}^{(i)}$ in (3.14) may be rewritten as

$$\begin{aligned} \hat{t}_{zc}^{(i)} &= \mathbf{t}_{\mathbf{x}}^{(i)\prime}\hat{\mathbf{B}}_{az}^{(i)} + \sum_{v=i}^{i+V-1}\sum\nolimits_{r_v^{(i)}} \frac{b_{vk}^{(i)} d_{vk}^{(i)}}{\pi_{vk}\hat{\theta}_{vk}^{(i)}} + \sum_{v=i}^{i}\sum\nolimits_{r_v^{(i)}} \frac{e_{vk}^{(i)}}{\pi_{vk}\hat{\theta}_{vk}^{(i)}} \\ &= \mathbf{t}_{\mathbf{x}}^{(i)\prime}\hat{\mathbf{B}}_{az}^{(i)} + \sum_{v=1}^{V^*}\sum\nolimits_{s_v} \frac{b_{vk}^{(i)} d_{vk}^{(i)} + e_{vk}^{(i)}}{\pi_{vk}\hat{\theta}_{vk}^{(i)}} \\ &= \mathbf{t}_{\mathbf{x}}^{(i)\prime}\hat{\mathbf{B}}_{az}^{(i)} + \sum_{v=1}^{V^*}\sum\nolimits_{s_v} \frac{R_{vk}^{(i)}(b_{vk}^{(i)} d_{vk}^{(i)} + e_{vk}^{(i)})}{\pi_{vk}\hat{\theta}_{vk}^{(i)}} \end{aligned}$$

The second stage, summing over s_v instead of $r_v^{(i)}$, is allowed, since for all combinations of $i = 1,...,I$ and $v = 1,...,V^*$, (a) the definition of $\hat{\theta}_{vk}^{(i)}$ is such that $\hat{\theta}_{vk}^{(i)} > 0$ for all $k \in s_v$ under assumptions *A1* and *A2*, and (b) $d_{vk}^{(i)}$ and $e_{vk}^{(i)}$ are defined for all $k \in s_v$. Consequently, an alternative expression for $\hat{T}_z$ is given by

$$
\begin{aligned}
\hat{T}_z &= \sum_{i=1}^{I} a^{(i)}[\mathbf{t}_{\mathbf{x}}^{(i)\prime}\hat{\mathbf{B}}_{az}^{(i)} + \sum_{v=1}^{V^*}\sum_{s_v} \frac{R_{vk}^{(i)}(b_{vk}^{(i)}d_{vk}^{(i)} + e_{vk}^{(i)})}{\pi_{vk}\hat{\theta}_{vk}^{(i)}}] \\
&= \sum_{i=1}^{I} a^{(i)}\mathbf{t}_{\mathbf{x}}^{(i)\prime}\hat{\mathbf{B}}_{az}^{(i)} + \sum_{i=1}^{I}\sum_{v=1}^{V^*}\sum_{s_v} \frac{a^{(i)}R_{vk}^{(i)}(b_{vk}^{(i)}d_{vk}^{(i)} + e_{vk}^{(i)})}{\pi_{vk}\hat{\theta}_{vk}^{(i)}} \\
&= \sum_{i=1}^{I} a^{(i)}\mathbf{t}_{\mathbf{x}}^{(i)\prime}\hat{\mathbf{B}}_{az}^{(i)} + \sum_{v=1}^{V^*}\sum_{s_v} \frac{\sum_{i=1}^{I} R_{vk}^{(i)}a^{(i)}(b_{vk}^{(i)}d_{vk}^{(i)} + e_{vk}^{(i)})/\hat{\theta}_{vk}^{(i)}}{\pi_{vk}} \\
&= \sum_{i=1}^{I} a^{(i)}\mathbf{t}_{\mathbf{x}}^{(i)\prime}\hat{\mathbf{B}}_{az}^{(i)} + \sum_{v=1}^{V^*}\sum_{s_v} \frac{\sum_{i=1}^{I} R_{vk}^{(i)}u_{vk}^{(i)}/\hat{\theta}_{vk}^{(i)}}{\pi_{vk}}
\end{aligned}
\tag{4.1}
$$

where

$$u_{vk}^{(i)} = a^{(i)}(b_{vk}^{(i)}d_{vk}^{(i)} + e_{vk}^{(i)})$$

For $i = 1,...,I$ and $v = 1,...,V^*$, define $d_{vk,U}^{(i)}$ and $e_{vk,U}^{(i)}$ for every $k \in U$, according to

$$d_{vk,U}^{(i)} = \begin{cases} \mathbf{y}_k^{(i)\prime}\mathbf{B}_z^{(i)} - \mathbf{x}_k^{(i)\prime}\mathbf{B}_{az}^{(i)} & \text{if } i \le v \le i + V - 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$e_{vk,U}^{(i)} = \begin{cases} Z_k^{(i)} - \mathbf{Y}_k^{(i)\prime}\mathbf{B}_z^{(i)} & \text{if } v = i \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbf{B}_{az}^{(i)} = E(\hat{\mathbf{B}}_{az}^{(i)})$ and $\mathbf{B}_z^{(i)} = E(\hat{\mathbf{B}}_z^{(i)})$, and let

$$u_{vk,U}^{(i)} = a^{(i)}(b_{vk}^{(i)}d_{vk,U}^{(i)} + e_{vk,U}^{(i)})$$

Substituting $\mathbf{B}_{az}^{(i)}$ and $u_{vk,U}^{(i)}$ for $\hat{\mathbf{B}}_{az}^{(i)}$ and $u_{vk}^{(i)}$, respectively, in (4.1) yields the approximation

$$\hat{T}_z \approx \sum_{i=1}^{I} a^{(i)}\mathbf{t}_{\mathbf{x}}^{(i)\prime}\mathbf{B}_{az}^{(i)} + \sum_{v=1}^{V^*}\sum_{s_v} \frac{\sum_{i=1}^{I} R_{vk}^{(i)}u_{vk,U}^{(i)}/\hat{\theta}_{vk}^{(i)}}{\pi_{vk}}$$

Hence, it follows that

$$V(\hat{T}_z) \approx V(\sum_{v=1}^{V^*}\sum\nolimits_{s_v} \frac{\sum_{i=1}^{I} R_{vk}^{(i)} u_{vk,U}^{(i)} / \hat{\theta}_{vk}^{(i)}}{\pi_{vk}}) = \sum_{v=1}^{V^*} V(\sum\nolimits_{s_v} \frac{\sum_{i=1}^{I} R_{vk}^{(i)} u_{vk,U}^{(i)} / \hat{\theta}_{vk}^{(i)}}{\pi_{vk}})$$

where the last step follows since (a) the samples $v = 1,...,V^*$ are drawn independently, and (b) any pair of elements k and l (including $k = l$) such that $k \in s_v$ and $l \in s_{v'}$, $v \neq v'$, respond independently at all times under the response distribution *RD*.

Since

$$E_{RD}(R_{vk}^{(i)} \mid \mathbf{m}_v^{(i)},...,\mathbf{m}_v^{(1)}, s_v) = \hat{\theta}_{vk}^{(i)} = \hat{\theta}_{vg}^{(i)} = m_{vg}^{(i)} / n_{vg}^{(i)}$$

under assumptions *A1* and *A2*, and $u_{vk,U}^{(i)}$ is as a non-random quantity for $k \in U$, $i = 1,...,I$ and $v = 1,...,V^*$, working with conditional expectations it follows that

$$E(\sum\nolimits_{s_v} \frac{\sum_{i=1}^{I} R_{vk}^{(i)} u_{vk,U}^{(i)} / \hat{\theta}_{vk}^{(i)}}{\pi_{vk}}) = \sum\nolimits_U \sum_{i=1}^{I} u_{vk,U}^{(i)}$$

Hence,

$$V(\sum\nolimits_{s_v} \frac{\sum_{i=1}^{I} R_{vk}^{(i)} u_{vk,U}^{(i)} / \hat{\theta}_{vk}^{(i)}}{\pi_{vk}}) = E[(\sum\nolimits_{s_v} \frac{\sum_{i=1}^{I} R_{vk}^{(i)} u_{vk,U}^{(i)} / \hat{\theta}_{vk}^{(i)}}{\pi_{vk}})^2] - (\sum\nolimits_U \sum_{i=1}^{I} u_{vk,U}^{(i)})^2 \tag{4.2}$$

Below, a variance estimator is proposed, using (4.2) as the starting point. For any set $S \subseteq U$, the notation $\sum\sum_S$ is shorthand for $\sum_{k \in S}\sum_{l \in S}$.

Suppose $u_{vk,U}^{(i)}$ could be observed for all $k \in r_v^{(i)}$, $i = 1,...,I$ and $v = 1,...,V^*$. In Appendix 3 it is shown that under assumptions *A1* and A2, an unbiased estimator for

$$(\sum\nolimits_U \sum_{i=1}^{I} u_{vk,U}^{(i)})^2 = (\sum_{i=1}^{I}\sum\nolimits_U u_{vk,U}^{(i)})^2 = \sum_{i=1}^{I}\sum_{i'=1}^{I}\sum\sum\nolimits_U u_{vk,U}^{(i)} u_{vl,U}^{(i')}$$

would be given by

$$\sum_{i=1}^{I}\sum_{i'=1}^{I}\sum\sum_{s_v}\frac{R_{vk}^{(i)}R_{vl}^{(i')}u_{vk,U}^{(i)}u_{vl,U}^{(i')}}{\pi_{vkl}\hat{\theta}_{vkl}^{(i,i')}} \tag{4.3}$$

where

$$\hat{\theta}_{vkl}^{(i,i')}=\begin{cases}\hat{\theta}_{vg}^{(i)}=m_{vg}^{(i)}/n_{vg}^{(i)} & \text{for } k=l,\ i=i',\ k\in s_{vg}^{(i)} \text{ and } g=1,...,G_v^{(i)}\\ \hat{\theta}_{vg}^{(i)}(m_{vg}^{(i)}-1)/(n_{vg}^{(i)}-1) & \text{for } k\neq l,\ i=i',\{k,l\}\in s_{vg}^{(i)} \text{ and } g=1,...,G_v^{(i)}\\ \hat{\theta}_{vg}^{(i)}\hat{\theta}_{vg'}^{(i')} & \text{for } k\in s_{vg}^{(i)} \text{ and } l\in s_{vg'}^{(i')} \text{ otherwise}\end{cases} \tag{4.4}$$

for $i=1,...,I$, $i'=1,...,I$ and $v=1,...,V^*$. Hence, if $u_{vk,U}^{(i)}$ could be observed

$$(\sum_{s_v}\frac{\sum_{i=1}^{I}R_{vk}^{(i)}u_{vk,U}^{(i)}/\hat{\theta}_{vk}^{(i)}}{\pi_{vk}})^2-\sum_{i=1}^{I}\sum_{i'=1}^{I}\sum\sum_{s_v}\frac{R_{vk}^{(i)}R_{vl}^{(i')}u_{vk,U}^{(i)}u_{vl,U}^{(i')}}{\pi_{vkl}\hat{\theta}_{vkl}^{(i,i')}} \tag{4.5}$$

would be unbiased for (4.2). Substituting $u_{vk}^{(i)}$ for $u_{vk,U}^{(i)}$ in (4.5) and summing over $v=1,...,V^*$, thus yields the working estimator

$$\hat{V}(\hat{T}_z)=\sum_{v=1}^{V^*}[(\sum_{s_v}\frac{\sum_{i=1}^{I}R_{vk}^{(i)}u_{vk}^{(i)}/\hat{\theta}_{vk}^{(i)}}{\pi_{vk}})^2-\sum_{i=1}^{I}\sum_{i'=1}^{I}\sum\sum_{s_v}\frac{R_{vk}^{(i)}R_{vl}^{(i')}u_{vk}^{(i)}u_{vl}^{(i')}}{\pi_{vkl}\hat{\theta}_{vkl}^{(i,i')}}] \tag{4.6}$$

with $\hat{\theta}_{vkl}^{(i,i')}$ defined according to (4.4). Using $z_k^{(i)}=y_{qk}^{(i)}$ in deriving $u_{vk}^{(i)}$, $i=1,...,I$, (4.6) becomes an estimator for $V(\hat{T}_{y_q})$, $q=1,...,Q$. The results of a small simulation study, mainly targeting the expected value of (4.6), are presented in Appendix 4.

4.2 Computational aspects

It is matter of algebra to show that $\hat{V}(\hat{T}_z)$ alternatively may be expressed as

$$\hat{V}(\hat{T}_z)=\sum_{v=1}^{V^*}\sum_{i=1}^{I}(\hat{V}_{1,r,v}^{(i)}+\hat{V}_{2,r,v}^{(i)})+\sum_{v=1}^{V^*}(\hat{V}_{s,v}-\sum_{i=1}^{I}\hat{V}_{s,v}^{(i)}) \tag{4.7}$$

where

$$\hat{V}_{1,r,v}^{(i)}=\sum\sum_{r_v^{(i)}}(\frac{\pi_{vkl}-\pi_{vk}\pi_{vl}}{\pi_{vkl}\hat{\theta}_{vkl}^{(i,i)}})\frac{u_{vk}^{(i)}}{\pi_{vk}}\frac{u_{vl}^{(i)}}{\pi_{vl}} \tag{4.8}$$

$$\hat{V}^{(i)}_{2,r,v} = \sum\sum_{r^{(i)}_v} \left(\frac{\hat{\theta}^{(i,i)}_{vkl} - \hat{\theta}^{(i)}_{vk}\hat{\theta}^{(i)}_{vl}}{\hat{\theta}^{(i,i)}_{vkl}}\right) \frac{u^{(i)}_{vk}}{\pi_{vk}\hat{\theta}^{(i)}_{vk}} \frac{u^{(i)}_{vl}}{\pi_{vl}\hat{\theta}^{(i)}_{vl}} \tag{4.9}$$

$$\hat{V}_{s,v} = \sum\sum_{s_v} \left(\frac{\pi_{vkl} - \pi_{vk}\pi_{vl}}{\pi_{vkl}}\right) \frac{\sum_{i=1}^{I} R^{(i)}_{vk} u^{(i)}_{vk} / \hat{\theta}^{(i)}_{vk}}{\pi_{vk}} \frac{\sum_{i=1}^{I} R^{(i)}_{vl} u^{(i)}_{vl} / \hat{\theta}^{(i)}_{vl}}{\pi_{vl}} \tag{4.10}$$

and

$$\hat{V}^{(i)}_{s,v} = \sum\sum_{s_v} \left(\frac{\pi_{vkl} - \pi_{vk}\pi_{vl}}{\pi_{vkl}}\right) \frac{R^{(i)}_{vk} u^{(i)}_{vk} / \hat{\theta}^{(i)}_{vk}}{\pi_{vk}} \frac{R^{(i)}_{vl} u^{(i)}_{vl} / \hat{\theta}^{(i)}_{vl}}{\pi_{vl}} \tag{4.11}$$

where $\hat{\theta}^{(i)}_{vk}$ and $\hat{\theta}^{(i,i)}_{vkl}$ are defined according to (4.4). Clearly, each of (4.8), (4.9), (4.10), and (4.11) algebraically resembles a variance estimator in its own respect. Hence, if software exists which allow for computation of (4.8), (4.9), (4.10), and (4.11), it may be used to compute $\hat{V}(\hat{T}_z)$ according to (4.7), given that the variables $u^{(i)}_{vk}$, $k \in r^{(i)}_v$, and $R^{(i)}_{vk} u^{(i)}_{vk} / \hat{\theta}^{(i)}_{vk}$, $k \in s_v$, can be derived for $i = 1,\ldots,I$ and $v = 1,\ldots,V^*$. Using $z^{(i)}_k = y^{(i)}_{qk}$ in deriving $u^{(i)}_{vk}$, $i = 1,\ldots,I$, (4.7) may also be used to compute $\hat{V}(\hat{T}_{y_q})$, $q = 1,\ldots,Q$.

5 An application - producing household-related statistics using the LFS

5.1 Sampling design in the LFS – an overview

Below, a brief overview of the LFS-design is given. The presentation only includes details deemed necessary for achieving a basic understanding of how household-related LFS statistics are produced. The presentation draws on the notation introduced in previous sections, but as the LFS-design is somewhat more intricate than the design introduced in section 2, some additional notation is introduced. For more information on the LFS-design, see Statistics Sweden (2011) and Statistics Sweden (2014b).

The LFS is conducted monthly. Calendar-months March, June, September and December consist of five reference weeks each, whereas the remaining months consist of four reference weeks[1]. Monthly estimates are combined to produce quarterly and annual estimates.

An individual included in the LFS is subject to data collection for a total of eight times, three months apart. In any given month, data are to be collected for a total of $V = 16$ subsamples. Each subsample is in itself a subsample from a larger sample, sometimes referred to as the annual sample. The annual sample for a specific year, say τ, includes all individuals that will enter the LFS for the first time during the year in question. As the LFS is conducted monthly, the set of individuals subject to data collection need to be refreshed each month. For this reason, the annual sample is split randomly into $I = 12$ subsamples, the i:th of which will be subject to data collection for the first time during month i in year τ. The rotational pattern implies the following:

- For January – September in reference year τ, the 16 subsamples used at the monthly level will be combination of subsamples from annual samples for years $\tau - 2$, $\tau - 1$ and τ.
- For October – December, the 16 subsamples used at the monthly level will be combination of subsamples from annual samples for years $\tau - 1$ and τ.

In any given month, the 16 subsamples can be divided into two groups of 8 subsamples each, after the stratification principles used when selecting the annual samples:

- For subsamples in the first group, the annual sample is drawn as a stratified systematic sample, using 48 strata defined after region and sex. This stratification will be referred to as stratification principle 1. Before the annual sample is drawn, individuals are ordered after country of birth and personal identification number within strata. In selecting the annual sample, four randomly chosen starting points are used.
- For subsamples in the second group, the annual sample is drawn as a stratified systematic sample, using 105 strata defined after country of birth,

[1] Every seventh year, the one additional month will consist of five reference weeks, thus making the LFS-year 53 weeks long.

age, region, and information from LISA and IoT, two registers held by Statistics Sweden. This stratification will be referred to as stratification principle 2. Before the annual sample is drawn, individuals are ordered after personal identification number within strata. In selecting the annual sample, four randomly chosen starting points are used.

In any given month i, $i = 1,...,I$, the following data are to be collected:

- $\mathbf{x}_k^{(i)} = (x_{1k}^{(i)}, x_{2k}^{(i)},..., x_{Pk}^{(i)})'$, the auxiliary vector used in the monthly LFS (for more information, see Statistics Sweden, 2011), and $\mathbf{y}_k^{(i)} = (y_{1k}^{(i)}, y_{2k}^{(i)},..., y_{Qk}^{(i)})'$, the vector of individual-level study-variables for which monthly statistics are to be produced and which are needed to fulfil the consistency requirement included in Commission Regulation (EC) No 430/2005, are to be collected for all individuals included in 16 subsamples subject to data collection.
- $\mathbf{Y}_k^{(i)}$ and $Z_k^{(i)}$ according to (2.3) and (2.4), respectively, with $c_k^{(i)} = 1/N_{j(k)}^{(i)}$ and clusters given by households, are to be collected for all individuals included in one subsample, namely the subsample drawn according to principle 1 which is subject to data collection for the eighth, and thus the last, time[2]. For Statistics Sweden, data collection on $\mathbf{Y}_k^{(i)}$ and $Z_k^{(i)}$ can be seen as a consequence of Regulation (EC) No 430/2005.

In principle, the LFS may be used to produce estimates of household-related parameters for the reference period month, but we will only consider the case when the reference period is year, i.e. a twelve-month period starting with January. The general parameter of interest is

$$T_z = \sum_{i=1}^{12} a^{(i)} t_z^{(i)} \tag{5.1}$$

with

$$a^{(i)} = \begin{cases} 5/52 & \text{for } i = 3,6,9,12 \\ 4/52 & \text{otherwise} \end{cases}$$

i.e., the *a*-weight is defined as the number of reference weeks in the month divided by the total number of reference weeks during the year.

From the description above, it follows that in order to produce an estimate of (5.1), a total of 66 subsamples are needed; 33 drawn according to stratification principle 1 and 33 drawn according to stratification principle 2. In Appendix 5, an overview of the subsample-structure and the data to be collected during each month is given. The presentation draws on the fact that since any given subsample included in the LFS will be subject to data collection at eight equidistant time periods, three

[2] Of course, more than one household related variable is to be collected, but for this report it is enough to consider the case when $Z^{(i)}$ is a scalar.

months apart, the 66 subsamples required to estimate (5.1) can be grouped into three groups such that each subsample belongs to one and only one subgroup.

Despite the somewhat complex sampling procedure described above, at the monthly level all subsamples are treated as independently drawn stratified simple random samples, with strata given by the strata used for drawing the annual samples. More formally, let $U_{\eta,v,h}$ denote the set of $N_{\eta,v,h}$ individuals that constituted stratum h in the annual sample, drawn according to stratification principle η, from which subsample $s_{\eta,v}$ was obtained. Furthermore, let $s_{\eta,v,h} = s_{\eta,v} \cap U_{\eta,v,h}$ denote the subset of $n_{\eta,v,h}$ individuals in $s_{\eta,v}$ that belong to stratum h, $\eta = 1,2$, $v = 1,...,33$, and $h = 1,..., H_\eta$. In producing LFS-statistics, the subsample $s_{\eta,v,h}$ is treated as a simple random sample from $U_{\eta,v,h}$. Thus, for $k \in s_{\eta,v,h}$, the first-order inclusion probability used is $\pi_{\eta,vk} = n_{\eta,v,h} / N_{\eta,v,h}$. For any pair of strata $h \neq h'$, $s_{\eta,v,h}$ and $s_{\eta,v,h'}$ are assumed to be drawn independently. From a statistical point of view, this implies ignoring the fact that subsamples obtained from the same annual sample are indeed independent. However, given the small sampling fractions, ignoring the dependency is merely a theoretical problem. Moreover, as the designs used for selecting the annual samples are likely to be at least as efficient as stratified simple random sampling using the same sample sizes, stratification-principles, and sample allocations, treating the subsamples as stratified simple random samples is expected to result in a conservative variance estimator.

5.2 Producing household-related LFS-statistics using HUUVA 1.0

As indicated in section 4.2, the software ETOS 2.0 may be used to produce an estimate of T_z in such a way that the consistency requirement Commission Regulation (EC) No 430/2005 is fulfilled. However, in order to simultaneously address the issues of point and variance estimation, a set of SAS-macros, collectively labelled HUUVA 1.0, has been derived. HUUVA 1.0 may be seen as an extension of ETOS 2.0, aimed at implementing the estimators presented in sections 3 and 4. However, in order to arrive at a working solution given the constraints imposed by the budget and time available for the task, the solution HUUVA 1.0 offers covers is somewhat restricted in comparison to the theoretical results presented in this report. Below, a brief description of how HUUVA 1.0 may be used for producing household-related LFS-statistics is provided. For more detailed information on HUUVA 1.0 in general, see Andersson (2015).

5.2.1 Point estimation

In principle, HUUVA 1.0 allows for estimation of (5.1) according to the estimator (3.13), the estimator derived in section 3.2, with response homogeneity groups coinciding with strata within subsamples. Let $r^{(i)}_{\eta,v,h} = r^{(i)}_{\eta,v} \cap U_{\eta,v,h}$ denote the subset of $m^{(i)}_{\eta,v,h}$ individuals in $r^{(i)}_{\eta,v}$ that constitute the response set at time i, for $\eta = 1,2$, $v = 1,...,33$, $h = 1,..., H_\eta$ and $i = 1,...,12$. Moreover, let $\Omega_i = \{i, i+3, i+6,..., i+21\}$, i.e., Ω_i denotes a set of numbers that indicate for what values of v, data for $k \in s_{\eta,v}$, $\eta = 1,2$, are to be collected at time i, $i = 1,...,12$. Taking the description given in section 5.1 and the presentation in section 3.2 into account, response homogeneity groups coinciding with strata within subsamples imply that for $k \in s_{\eta,v,h}$, $\eta = 1,2$, $v = 1,...,33$, $h = 1,..., H_\eta$,

$$\hat{\theta}^{(i)}_{\eta,vk} = \begin{cases} m^{(i)}_{\eta,v,h} / n_{\eta,v,h} & \text{if } v \in \Omega_i \\ n_{\eta,v,h} / n_{\eta,v,h} & \text{otherwise} \end{cases}$$

However, in order to obtain a weight-system that exactly fulfills the consistency requirement, some minor adaption of (3.13) is necessary. The starting point is the following expression from (3.14),

$$\hat{t}^{(i)}_{zc} = (\hat{\mathbf{t}}^{(i)}_{ayc} - \hat{\mathbf{t}}^{(i)}_{\mathbf{Y}})'\hat{\mathbf{B}}^{(i)}_z + \hat{t}^{(i)}_Z \qquad (5.2)$$

Rather than using (3.13) with (5.2) as the estimator for $t^{(i)}_{zc}$, household-estimates are based on (5.2) with $\hat{\mathbf{t}}^{(i)}_{ayc}$ replaced by $\hat{\mathbf{t}}^{(i)}_{ayc,LFS}$, the official monthly LFS-estimate of $\mathbf{t}^{(i)}_{\mathbf{y}}$ for month $i = 1,...,12$. The main reasons for this is that at time i, $i = 1,...,12$, HUUVA 1.0 will use only those individuals for which data are available on both the individual-level variables $\mathbf{x}^{(i)}_k$ and $\mathbf{y}^{(i)}_k$, and the household-associated variables $\mathbf{Y}^{(i)}_k$ and $Z^{(i)}_k$, in defining the response set $r^{(i)}_v$ for $v = i$. However, in the production of official monthly LFS-statistics, for month $i = 1,...,12$, all response sets are defined on the basis of availability to data on the individual-level variables $\mathbf{x}^{(i)}_k$ and $\mathbf{y}^{(i)}_k$ only. Hence, in order to obtain a weight-system which is in line with the requirements, using $\hat{\mathbf{t}}^{(i)}_{ayc,LFS}$ is essential. Moreover, using $\hat{\mathbf{t}}^{(i)}_{ayc,LFS}$ in deriving the estimator for T_z, rather than computing $\hat{\mathbf{t}}^{(i)}_{ayc}$, according to (3.12), from the set of microdata provided as input to HUUVA 1.0, means using more of the data actually available.

To compensate for the fact that the number of respondents who provide household data at the monthly level is fairly small, given the large number of classes implied by the vectors $\mathbf{x}$ and $\mathbf{y}$ used in the estimation, regression estimation is applied at the quarterly level rather than the monthly. The starting point is the following alternative expression for (5.1),

$$T_z = \frac{\sum_{\kappa=1}^{4} t^{(\kappa)}_z}{4}$$

where

$$t^{(\kappa)}_z = \sum_{i=3(\kappa-1)+1}^{3\kappa} 4a^{(i)} t^{(i)}_z \qquad (5.3)$$

with $\kappa = 1,...,4$ indicating quarter. That is, T_z may be seen as an average of four quarterly totals, where each quarterly total in itself is a weighted average of the monthly totals for the months in the quarter. The factor four in (5.3) serves to rescale the a-weights to the quarterly level, i.e., to guarantee that the weight system used sums to one for each quarter.

For $\kappa = 1,...,4$, let

$$\hat{\mathbf{t}}^{(\kappa)}_{ayc,LFS} = \sum_{i=3(\kappa-1)+1}^{3\kappa} 4a^{(i)}\hat{\mathbf{t}}^{(i)}_{ayc,LFS}$$

Moreover, let

$$\hat{\mathbf{t}}^{(\kappa)}_{\mathbf{Y}} = \sum_{i=3(\kappa-1)+1}^{3\kappa} 4a^{(i)}\hat{\mathbf{t}}^{(i)}_{\mathbf{Y}}$$

with

$$\hat{\mathbf{t}}^{(i)}_{\mathbf{Y}} = \sum_{v=i}^{i}\sum\nolimits_{r^{(i)}_{1,v}} \frac{\mathbf{Y}^{(i)}_k}{\pi_{1,vk}\hat{\theta}^{(i)}_{1,vk}} = \sum_{v=i}^{i}\sum_{h=1}^{H_1} \frac{N_{1,v,h}}{m^{(i)}_{1,v,h}} \sum\nolimits_{r^{(i)}_{1,v,h}} \mathbf{Y}^{(i)}_k$$

$$\hat{\mathbf{B}}^{(\kappa)}_z = (\sum_{i=3(\kappa-1)+1}^{3\kappa} 4a^{(i)}\hat{\mathbf{T}}^{(i)}_{\mathbf{YY}})^{-1} \sum_{i=3(\kappa-1)+1}^{3\kappa} 4a^{(i)}\hat{\mathbf{T}}^{(i)}_{\mathbf{YZ}}$$

with

$$\hat{\mathbf{T}}^{(i)}_{\mathbf{YY}} = \sum_{v=i}^{i}\sum\nolimits_{r^{(i)}_{1,v}} \frac{\mathbf{Y}^{(i)}_k\mathbf{Y}^{(i)\prime}_k}{\pi_{1,vk}\hat{\theta}^{(i)}_{1,vk}} = \sum_{v=i}^{i}\sum_{h=1}^{H_1} \frac{N_{1,v,h}}{m^{(i)}_{1,v,h}} \sum\nolimits_{r^{(i)}_{1,v,h}} \mathbf{Y}^{(i)}_k\mathbf{Y}^{(i)\prime}_k$$

and

$$\hat{\mathbf{T}}^{(i)}_{\mathbf{Y}Z} = \sum_{v=i}^{i}\sum\nolimits_{r^{(i)}_{1,v}} \frac{\mathbf{Y}^{(i)}_k Z^{(i)}_k}{\pi_{1,vk}\hat{\theta}^{(i)}_{1,vk}} = \sum_{v=i}^{i}\sum_{h=1}^{H_1} \frac{N_{1,v,h}}{m^{(i)}_{1,v,h}} \sum\nolimits_{r^{(i)}_{1,v,h}} \mathbf{Y}^{(i)}_k Z^{(i)}_k$$

respectively, and

$$\hat{t}^{(\kappa)}_Z = \sum_{i=3(\kappa-1)+1}^{3\kappa} 4a^{(i)}\hat{t}^{(i)}_Z$$

with

$$\hat{t}_Z^{(i)} = \sum_{v=i}^{i} \sum_{r_{1,v}^{(i)}} \frac{Z_k^{(i)}}{\pi_{1,vk}\hat{\theta}_{1,vk}^{(i)}} = \sum_{v=i}^{i} \sum_{h=1}^{H_1} \frac{N_{1,v,h}}{m_{1,v,h}^{(i)}} \sum_{r_{1,v,h}^{(i)}} Z_k^{(i)}$$

The estimator actually used for point estimation of household-related statistics is

$$\tilde{T}_z = \frac{\sum_{\kappa=1}^{4} \tilde{t}_{zc}^{(\kappa)}}{4} \tag{5.4}$$

with

$$\tilde{t}_{zc}^{(\kappa)} = (\hat{\mathbf{t}}_{ayc,LFS}^{(\kappa)} - \hat{\mathbf{t}}_{\mathbf{Y}}^{(\kappa)})'\hat{\mathbf{B}}_z^{(\kappa)} + \hat{t}_Z^{(\kappa)}$$

Whereas $\hat{\mathbf{t}}_{ayc,LFS}^{(\kappa)}$ must be provided as part of the input to HUUVA 1.0, $\hat{\mathbf{t}}_{\mathbf{Y}}^{(\kappa)}$, $\hat{\mathbf{B}}_z^{(\kappa)}$, and $\hat{t}_Z^{(\kappa)}$, $\kappa = 1,...,4$, are computed directly from the microdata which must be part of the input to HUUVA 1.0.

5.2.2 Variance estimation

For $k \in s_{\eta,v}$, let $b_{\eta,vk}^{(i)}$ denote the element-specific weight associated with element k in the production of LFS-statistics for month $i = 1,...,12$. The results in sections 3 and 4 are based on the explicit assumption that the b-weights are non-stochastic. However, in the LFS, this assumption is not fulfilled. In essence, at time i, $b_{\eta,vk}^{(i)}$ for $k \in s_{\eta,v}$, $v \in \Omega_i$, may be interpreted as a nonresponse adjusted version of the weight

$$\frac{\pi_{\eta,vk}}{\sum_{\eta=1}^{2} \sum_{v' \in V_i} \pi_{\eta,v'k}} \approx \frac{\pi_{\eta,vk}}{8(\pi_{1,vk} + \pi_{2,vk})}$$

Nevertheless, the b-weights are treated as non-stochastic in the monthly estimation.

Although it is highly likely that for any individual included in the LFS, the b-weights will vary somewhat over time, HUUVA 1.0 only allows for one b-weight per sampling unit. Therefore,

$$b_{\eta,vk} = \frac{\sum_{i=1}^{12} I(v \in \Omega_i) R_{\eta,vk}^{(i)} b_{\eta,vk}^{(i)}}{\sum_{i=1}^{12} I(v \in \Omega_i) R_{\eta,vk}^{(i)}}$$

where

$$I(v \in \Omega_i) = \begin{cases} 1 & \text{if } v \in \Omega_i \\ 0 & \text{otherwise} \end{cases}$$

and

$$R_{\eta,vk}^{(i)} = \begin{cases} 0 & \text{if } k \notin r_{\eta,v}^{(i)}, i \le v \le i+V-1 \\ 1 & \text{otherwise} \end{cases}$$

$\eta = 1,2$, $v = 1,...,33$, $i = 1,...,12$, is used when producing variance estimates for household-related statistics. In analogy with the monthly LFS-production, the b -weights are treated as non-stochastic in the estimation.

For $\kappa = 1,...,4$, let

$$\mathbf{t}_{\mathbf{x}}^{(\kappa)} = \sum_{i=3(\kappa-1)+1}^{3\kappa} 4a^{(i)} \mathbf{t}_{\mathbf{x}}^{(i)}$$

$$\hat{\mathbf{t}}_{a\mathbf{x}}^{(\kappa)} = \sum_{i=3(\kappa-1)+1}^{3\kappa} 4a^{(i)} \hat{\mathbf{t}}_{a\mathbf{x}}^{(i)}$$

with

$$\hat{\mathbf{t}}_{a\mathbf{x}}^{(i)} = \sum_{\eta=1}^{2} \sum_{v\in\Omega_i} \sum\nolimits_{r_{\eta,v}^{(i)}} \frac{b_{\eta,vk}\mathbf{x}_k^{(i)}}{\pi_{\eta,vk}\hat{\theta}_{\eta,vk}^{(i)}} = \sum_{\eta=1}^{2} \sum_{v\in\Omega_i} \sum_{h=1}^{H_\eta} \frac{N_{\eta,v,h}}{m_{\eta,v,h}^{(i)}} \sum\nolimits_{r_{\eta,v,h}^{(i)}} b_{\eta,vk}\mathbf{x}_k^{(i)}$$

$$\hat{\mathbf{t}}_{a\mathbf{y}}^{(\kappa)} = \sum_{i=3(\kappa-1)+1}^{3\kappa} 4a^{(i)} \hat{\mathbf{t}}_{a\mathbf{y}}^{(i)}$$

with $\hat{\mathbf{t}}_{a\mathbf{y}}^{(i)}$ analogous to $\hat{\mathbf{t}}_{a\mathbf{x}}^{(i)}$, and

$$\hat{\mathbf{B}}_{a\mathbf{y}}^{(\kappa)} = \left(\sum_{i=3(\kappa-1)+1}^{3\kappa} 4a^{(i)} \hat{\mathbf{T}}_{\mathbf{xx}}^{(i)} \right)^{-1} \sum_{i=3(\kappa-1)+1}^{3\kappa} 4a^{(i)} \hat{\mathbf{T}}_{\mathbf{xy}}^{(i)}$$

with

$$\hat{\mathbf{T}}_{\mathbf{xx}}^{(i)} = \sum_{\eta=1}^{2} \sum_{v\in\Omega_i} \sum\nolimits_{r_{\eta,v}^{(i)}} \frac{b_{\eta,vk}\mathbf{x}_k^{(i)}\mathbf{x}_k^{(i)\prime}}{\pi_{\eta,vk}\hat{\theta}_{\eta,vk}^{(i)}} = \sum_{\eta=1}^{2} \sum_{v\in\Omega_i} \sum_{h=1}^{H_\eta} \frac{N_{\eta,v,h}}{m_{\eta,v,h}^{(i)}} \sum\nolimits_{r_{\eta,v,h}^{(i)}} b_{\eta,vk}\mathbf{x}_k^{(i)}\mathbf{x}_k^{(i)\prime}$$

and

$$\hat{\mathbf{T}}_{\mathbf{xy}}^{(i)} = \sum_{\eta=1}^{2}\sum_{v\in\Omega_i}\sum\nolimits_{r_{\eta,v}^{(i)}} \frac{b_{\eta,vk}\mathbf{x}_k^{(i)}\mathbf{y}_k^{(i)\prime}}{\pi_{\eta,vk}\hat{\theta}_{\eta,vk}^{(i)}} = \sum_{\eta=1}^{2}\sum_{v\in V_i}\sum_{h=1}^{H_\eta} \frac{N_{\eta,v,h}}{m_{\eta,v,h}^{(i)}} \sum\nolimits_{r_{\eta,v,h}^{(i)}} b_{\eta,vk}\mathbf{x}_k^{(i)}\mathbf{y}_k^{(i)\prime}$$

respectively. Using the above definitions, together with $\hat{\mathbf{t}}_{\mathbf{Y}}^{(\kappa)}$, $\hat{\mathbf{B}}_z^{(\kappa)}$, and $\hat{t}_Z^{(\kappa)}$ as defined in section 5.2.1, let

$$\hat{T}_z = \frac{\sum_{\kappa=1}^{4}\hat{t}_{zc}^{(\kappa)}}{4} \tag{5.5}$$

where

$$\hat{t}_{zc}^{(\kappa)} = (\mathbf{t}_{\mathbf{x}}^{(\kappa)} - \hat{\mathbf{t}}_{a\mathbf{x}}^{(\kappa)})'\hat{\mathbf{B}}_{az}^{(\kappa)} + (\hat{\mathbf{t}}_{a\mathbf{y}}^{(\kappa)} - \hat{\mathbf{t}}_{\mathbf{Y}}^{(\kappa)})'\hat{\mathbf{B}}_z^{(\kappa)} + \hat{t}_Z^{(\kappa)} \tag{5.6}$$

with $\hat{\mathbf{B}}_{az}^{(\kappa)} = \hat{\mathbf{B}}_{a\mathbf{y}}^{(\kappa)}\hat{\mathbf{B}}_z^{(\kappa)}$. Even though point estimates are derived using $\tilde{T}_z$ according to (5.4), variance estimates are derived under the assumption that the point estimator used is $\hat{T}_z$ according to (5.5), which can be computed directly from the set of microdata which must be part of the input to HUUVA 1.0.

The variance estimates are computed according to (4.7), taking into account the assumptions and modifications previously introduced in section 5. For $k \in s_{\eta,v}$, let

$$u_{\eta,vk}^{(i)} = a^{(i)}(b_{\eta,vk}d_{\eta,vk}^{(i)} + e_{\eta,vk}^{(i)})$$

with

$$d_{\eta,vk}^{(i)} = \begin{cases} \mathbf{y}_k^{(i)\prime}\hat{\mathbf{B}}_z^{(\kappa(i))} - \mathbf{x}_k^{(i)\prime}\hat{\mathbf{B}}_{az}^{(\kappa(i))} & \text{if } k \in r_{\eta,v}^{(i)} \text{ and } v \in V_i \\ 0 & \text{otherwise} \end{cases}$$

and

$$e_{\eta,vk}^{(i)} = \begin{cases} Z_k^{(i)} - \mathbf{Y}_k^{(i)\prime}\hat{\mathbf{B}}_z^{(\kappa(i))} & \text{if } k \in r_{1,v}^{(i)} \text{ and } v = i \\ 0 & \text{otherwise} \end{cases}$$

where $\kappa(i)$ denotes the quarter to which month i belongs, $\eta = 1,2$, $v = 1,\ldots,33$, $i = 1,\ldots,12$. After proper modification, the variance estimator actually used becomes

$$\hat{V}(\hat{T}_z) = \sum_{\eta=1}^{2}\sum_{v=1}^{33}\sum_{i=1}^{12}(\hat{V}_{1,\eta,r,v}^{(i)} + \hat{V}_{2,\eta,r,v}^{(i)}) + \sum_{\eta=1}^{2}\sum_{v=1}^{33}(\hat{V}_{\eta,s,v} - \sum_{i=1}^{12}\hat{V}_{\eta,s,v}^{(i)}) \qquad (5.7)$$

where

$$\hat{V}_{1,\eta,r,v}^{(i)} = \sum_{h=1}^{H_v}\frac{N_{\eta,v,h}^2}{n_{\eta,v,h}}(1-\frac{n_{\eta,v,h}}{N_{\eta,v,h}})\frac{\sum_{r_{\eta,v,h}^{(i)}}(u_{\eta,vk}^{(i)})^2 - (\sum_{r_{\eta,v,h}^{(i)}}u_{\eta,vk}^{(i)})^2 / m_{\eta,v,h}^{(i)}}{m_{\eta,v,h}^{(i)} - 1}$$

$$\hat{V}_{2,\eta,r,v}^{(i)} = \sum_{h=1}^{H_v}\frac{N_{\eta,v,h}^2}{m_{\eta,v,h}^{(i)}}(1-\frac{m_{\eta,v,h}^{(i)}}{n_{\eta,v,h}})\frac{\sum_{r_{\eta,v,h}^{(i)}}(u_{\eta,vk}^{(i)})^2 - (\sum_{r_{\eta,v,h}^{(i)}}u_{\eta,vk}^{(i)})^2 / m_{\eta,v,h}^{(i)}}{m_{\eta,v,h}^{(i)} - 1}$$

$$\hat{V}_{\eta,s,v} = \sum_{h=1}^{H_v}\frac{N_{\eta,v,h}^2}{n_{\eta,v,h}}(1-\frac{n_{\eta,v,h}}{N_{\eta,v,h}})\frac{\sum_{s_{\eta,v,h}}(\sum_{i=1}^{12}R_{\eta,vk}^{(i)}u_{\eta,vk}^{(i)} / \hat{\theta}_{\eta,vk}^{(i)})^2 - (\sum_{s_{\eta,v,h}}\sum_{i=1}^{12}R_{\eta,vk}^{(i)}u_{\eta,vk}^{(i)} / \hat{\theta}_{\eta,vk}^{(i)})^2 / n_{\eta,v,h}}{n_{\eta,v,h} - 1}$$

and

$$\hat{V}_{\eta,s,v}^{(i)} = \sum_{h=1}^{H_v}\frac{N_{\eta,v,h}^2}{n_{\eta,v,h}}(1-\frac{n_{\eta,v,h}}{N_{\eta,v,h}})\frac{\sum_{s_{\eta,v,h}}(R_{\eta,vk}^{(i)}u_{\eta,vk}^{(i)} / \hat{\theta}_{\eta,vk}^{(i)})^2 - (\sum_{s_{\eta,v,h}}R_{\eta,vk}^{(i)}u_{\eta,vk}^{(i)} / \hat{\theta}_{\eta,vk}^{(i)})^2 / n_{\eta,v,h}}{n_{\eta,v,h} - 1}$$

Thus, the variance estimate will reflect the variance of $\hat{T}_z$ according to (5.5), rather than the variance of the used point estimator, $\tilde{T}_z$ according to (5.4). However, the fact that $\tilde{T}_z$ and $\hat{T}_z$ are defined in the same manner, the main difference being that $\tilde{T}_z$ is based on somewhat more available data than $\hat{T}_z$, suggests that $V(\tilde{T}_z) \le V(\hat{T}_z)$. Under assumptions *A1* and *A2* made in section 3.2, using $\hat{V}(\hat{T}_z)$ according to (5.7) as the estimator for $V(\tilde{T}_z)$, should thus amount to using a conservative variance estimator.

It should be noted that for $\eta = 1,2$ and $i = 1,...,12$,

$$\begin{aligned}\hat{V}_{1,\eta,r,v}^{(i)} + \hat{V}_{2,\eta,r,v}^{(i)} &= \sum_{h=1}^{H_v}\frac{N_{\eta,v,h}^2}{m_{\eta,v,h}^{(i)}}(1-\frac{m_{\eta,v,h}^{(i)}}{N_{\eta,v,h}})\frac{\sum_{r_{\eta,v,h}^{(i)}}(u_{\eta,vk}^{(i)})^2 - (\sum_{r_{\eta,v,h}^{(i)}}u_{\eta,vk}^{(i)})^2 / m_{\eta,v,h}^{(i)}}{m_{\eta,v,h}^{(i)} - 1} \\ &\approx \sum_{h=1}^{H_v}(\frac{N_{\eta,v,h}}{m_{\eta,v,h}^{(i)}})^2[\sum_{r_{\eta,v,h}^{(i)}}(u_{\eta,vk}^{(i)})^2 - \frac{(\sum_{r_{\eta,v,h}^{(i)}}u_{\eta,vk}^{(i)})^2}{m_{\eta,v,h}^{(i)}}]\end{aligned}$$

for $v \in \Omega_i$, and $\hat{V}_{1,\eta,r,v}^{(i)} + \hat{V}_{2,\eta,r,v}^{(i)} = 0$ for $v \notin \Omega_i$. Moreover, as $\hat{\theta}_{\eta,vk}^{(i)} = m_{\eta,v,h}^{(i)} / n_{\eta,v,h}$ for $k \in r_{\eta,v,h}^{(i)}$, $\eta = 1,2$, $v = 1,...,33$ and $i = 1,...,12$, it follows that

$$\hat{V}^{(i)}_{\eta,s,v} = (\frac{n_{\eta,v,h}}{m^{(i)}_{\eta,v,h}})^2 \frac{N^2_{\eta,v,h}}{n_{\eta,v,h}}(1-\frac{n_{\eta,v,h}}{N_{\eta,v,h}})\frac{\sum_{r^{(i)}_{\eta,v,h}}(u^{(i)}_{\eta,vk})^2 - (\sum_{r^{(i)}_{\eta,v,h}} u^{(i)}_{\eta,vk})^2 / n_{\eta,v,h}}{n_{\eta,v,h}-1}$$

$$\approx (\frac{N_{\eta,v,h}}{m^{(i)}_{\eta,v,h}})^2 [\sum_{r^{(i)}_{\eta,v,h}}(u^{(i)}_{\eta,vk})^2 - \frac{(\sum_{r^{(i)}_{\eta,v,h}} u^{(i)}_{\eta,vk})^2}{n_{\eta,v,h}}]$$

for $v \in \Omega_i$, and $\hat{V}^{(i)}_{\eta,s,v} = 0$ for $v \notin \Omega_i$. Hence

$$\hat{V}(\hat{T}_z) = \sum_{\eta=1}^{2}\sum_{v=1}^{33}\sum_{i=1}^{12}(\hat{V}^{(i)}_{1,\eta,r,v} + \hat{V}^{(i)}_{2,\eta,r,v}) + \sum_{\eta=1}^{2}\sum_{v=1}^{33}(\hat{V}_{\eta,s,v} - \sum_{i=1}^{12}\hat{V}^{(i)}_{\eta,s,v})$$

$$= \sum_{\eta=1}^{2}\sum_{v=1}^{33}\hat{V}_{\eta,s,v} + \sum_{\eta=1}^{2}\sum_{v=1}^{33}\sum_{i=1}^{12}(\hat{V}^{(i)}_{1,\eta,r,v} + \hat{V}^{(i)}_{2,\eta,r,v} - \hat{V}^{(i)}_{\eta,s,v})$$

$$\approx \sum_{\eta=1}^{2}\sum_{v=1}^{33}\hat{V}_{\eta,s,v} + \sum_{\eta=1}^{2}\sum_{v=1}^{33}\sum_{i=1}^{12}(\frac{N_{\eta,v,h}}{m^{(i)}_{\eta,v,h}})^2(\sum_{r^{(i)}_{\eta,v,h}} u^{(i)}_{\eta,vk})^2(\frac{1}{n_{\eta,v,h}} - \frac{1}{m^{(i)}_{\eta,v,h}})$$

Thus, if $(\sum_{r^{(i)}_{\eta,v,h}} u^{(i)}_{\eta,vk})^2 \approx 0$ for all $\eta = 1,2$, $v = 1,\ldots,33$ and $i = 1,\ldots,12$, it follows that

$$\hat{V}(\hat{T}_z) \approx \sum_{\eta=1}^{2}\sum_{v=1}^{33}\hat{V}_{\eta,s,v}$$

which is very easy to compute.

6 Concluding remarks

Design and analysis of surveys over time is an issue of great importance in the production of official statistics. Examples of relevant, fairly recent, references are Duncan and Kalton (1987), Binder (1998), Kish (1998), Holmes and Skinner (2000), Nordberg (2000), Berger (2006), and Steel and McLaren (2008). Even so, there are outstanding issues, in particular regarding variance estimation. The main reason for this is that the issue quickly becomes complex, when the specificities of sampling design, nonresponse and point estimation are taken into account.

The work presented in this report emanates from work carried out regarding estimation of household-related statistics in the LFS. More specifically, at the outset the goal was to decide on how to compute point and variance estimates in such a way that (a) the used estimators are theoretically justified and (b) the consistency constraint imposed by Commission Regulation (EC) No 430/2005 is fulfilled. As a bonus, the proposed solution turned out to be possible to implement through adaptation of the already existing software ETOS 2.0, the result being the software HUUVA 1.0. Since the software was developed with implementation for the LFS in mind, it is somewhat less general than the theory presented in sections 3 and 4. On the other hand, HUUVA 1.0 is very general in terms of what parameters it provides point and variance estimates for. Although the focus in the report has been on estimation of T_z, the software allows for point and variance estimates to be derived for any parameter of the type

$$\Theta = f(\mathbf{T_z})$$

where f denotes a rational, scalar-valued, function and $\mathbf{T_z}$ is a vector of household-related totals.

Despite being initially derived with the LFS in mind, the general results presented in this report are valid for any sample survey that can be described using the generic design introduced in section 2. Moreover, the results presented can easily be extended to cover also point and variance estimation for estimators of change over time. Thus, the theoretical results presented should prove relevant and useful in a more general context as well.

References

Andersson, C. (2012). ETOS 2.0 User's guide. Statistics Sweden.

Andersson, C. (2015). HUUVA 1.0. Statistics Sweden.

Axelson, M. (2000). A note on variance estimation for the regression estimator under two-phase sampling. In *On variance estimation for the two-phase regression estimator*. Ph. D. thesis, Department of Statistics, Uppsala University.

Berger, Y. G. (2004). *Variance estimation for measures of change in probability sampling*. Canadian Journal of Statistics 32, pp. 451–467.

Binder, D. A. (1998). Longitudinal surveys: why are these surveys different from all other surveys? *Survey Methodology* 24, pp. 101-108.

Duncan, G. J., and Kalton, G. (1987). Issues of design and analysis of surveys across time. *International Statistical Review* 55, pp. 97-117.

Holmes, D.J., and Skinner, C.J. (2000). Variance Estimation for Labour Force Survey Estimates of Level and Change. Government Statistical Service Methodology Series 21, United Kingdom. (http://www.ons.gov.uk/ons/guide-method/method-quality/specific/gss-methodology-series/gss-methodology-series--21--variation-estimation-for-labour-force-survey-estimates-of-level-and-change.pdf)

International Labour Organization (2013). *Resolution concerning statistics of work, employment and labour underutilization*. (http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_230304.pdf)

Kish, L. (1998). *Space / time variations and rolling samples*. Journal of Official Statistics 14, pp. 31-46.

Statistics Sweden (2011). The Sampling and Estimation Procedure in the Swedish Labour Force Surveys 2005-. Background Facts, Labour and Education Statistics 2011:6. Statistics Sweden. (http://www.scb.se/statistik/_publikationer/AM0401_2011A01_BR_AM76BR1106.pdf)

Statistics Sweden (2014a). Method for estimation when combining samples with different designs in the Swedish Labour Force Surveys. Background Facts, Labour and Education Statistics 2014:1, Statistics Sweden. (http://www.scb.se/Statistik/_Publikationer/AM0401_2014M01_BR_AM76BR1401.pdf)

Statistics Sweden (2014b). The Labour Force Surveys 2014. Description of the statistics, AM0401, Statistics Sweden. (http://www.scb.se/Statistik/AM/AM0401/_dokument/AM0401_BS_2014_eng.pdf)

Steel, D. and McLaren, C. (2009). *Design and analysis of surveys repeated over time*. Handbook of Survey Statistics 29B, Amsterdam: Elsevier, pp. 289 – 313.

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

Vale, C. D. and Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika* 48, pp. 465-471.

Appendix 1 Data to be collected, $I = 3$ and $V = 3$

For $I = 3$ and $V = 3$, the notation in section 2 implies that the following data are to be collected:

For	**Time-point *i***		
	1	**2**	**3**
$k \in s_1$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$, $\mathbf{Y}_k^{(1)}$ and $Z_k^{(1)}$		
$k \in s_2$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$, $\mathbf{Y}_k^{(2)}$ and $Z_k^{(2)}$	
$k \in s_3$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$, $\mathbf{Y}_k^{(3)}$ and $Z_k^{(3)}$
$k \in s_4$		$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$
$k \in s_5$			$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$

Appendix 2 Definitions of $d_{vk}^{(i)}$ and $e_{vk}^{(i)}$, $I = 3$ and $V = 3$

For $I = 3$ and $V = 3$, the notation in section 4 implies the following definitions for $d_{vk}^{(i)}$ and $e_{vk}^{(i)}$:

v	**Time-point *i***		
	1	**2**	**3**
		For $k \in s_v$	
1	For $k \in r_1^{(1)}$: $d_{1k}^{(1)} = \mathbf{y}_k^{(1)\prime}\hat{\mathbf{B}}_z^{(1)} - \mathbf{x}_k^{(1)\prime}\hat{\mathbf{B}}_{az}^{(1)}$ $e_{1k}^{(1)} = Z_k^{(1)} - \mathbf{Y}_k^{(1)\prime}\hat{\mathbf{B}}_z^{(1)}$ For $k \notin r_1^{(1)}$: $d_{1k}^{(1)} = 0$, $e_{1k}^{(1)} = 0$	For $k \in r_1^{(2)} (= s_1)$: $d_{1k}^{(2)} = 0$, $e_{1k}^{(2)} = 0$	For $k \in r_1^{(3)} (= s_1)$: $d_{1k}^{(3)} = 0$. $e_{1k}^{(3)} = 0$
2	For $k \in r_2^{(1)}$: $d_{2k}^{(1)} = \mathbf{y}_k^{(1)\prime}\hat{\mathbf{B}}_z^{(1)} - \mathbf{x}_k^{(1)\prime}\hat{\mathbf{B}}_{az}^{(1)}$ $e_{2k}^{(1)} = 0$ For $k \notin r_2^{(1)}$: $d_{2k}^{(1)} = 0$, $e_{2k}^{(1)} = 0$	For $k \in r_2^{(2)}$: $d_{2k}^{(2)} = \mathbf{y}_k^{(2)\prime}\hat{\mathbf{B}}_z^{(2)} - \mathbf{x}_k^{(2)\prime}\hat{\mathbf{B}}_{az}^{(2)}$ $e_{2k}^{(2)} = Z_k^{(2)} - \mathbf{Y}_k^{(2)\prime}\hat{\mathbf{B}}_z^{(2)}$ For $k \notin r_2^{(2)}$: $d_{2k}^{(2)} = 0$, $e_{2k}^{(2)} = 0$	For $k \in r_2^{(3)} (= s_2)$: $d_{2k}^{(3)} = 0$, $e_{2k}^{(3)} = 0$
3	For $k \in r_3^{(1)}$: $d_{3k}^{(1)} = \mathbf{y}_k^{(1)\prime}\hat{\mathbf{B}}_z^{(1)} - \mathbf{x}_k^{(1)\prime}\hat{\mathbf{B}}_{az}^{(1)}$ $e_{3k}^{(1)} = 0$ For $k \notin r_3^{(1)}$: $d_{3k}^{(1)} = 0$, $e_{3k}^{(1)} = 0$	For $k \in r_3^{(2)}$: $d_{3k}^{(2)} = \mathbf{y}_k^{(2)\prime}\hat{\mathbf{B}}_z^{(2)} - \mathbf{x}_k^{(2)\prime}\hat{\mathbf{B}}_{az}^{(2)}$ $e_{3k}^{(2)} = 0$ For $k \notin r_3^{(2)}$: $d_{3k}^{(2)} = 0$, $e_{3k}^{(2)} = 0$	For $k \in r_3^{(3)}$: $d_{3k}^{(3)} = \mathbf{y}_k^{(3)\prime}\hat{\mathbf{B}}_z^{(3)} - \mathbf{x}_k^{(3)\prime}\hat{\mathbf{B}}_{az}^{(3)}$ $e_{3k}^{(3)} = Z_k^{(3)} - \mathbf{Y}_k^{(3)\prime}\hat{\mathbf{B}}_z^{(3)}$ For $k \notin r_3^{(3)}$: $d_{3k}^{(3)} = 0$, $e_{3k}^{(3)} = 0$
4	For $k \in r_4^{(1)} (= s_4)$: $d_{4k}^{(1)} = 0$, $e_{4k}^{(1)} = 0$	For $k \in r_4^{(2)}$: $d_{4k}^{(2)} = \mathbf{y}_k^{(2)\prime}\hat{\mathbf{B}}_z^{(2)} - \mathbf{x}_k^{(2)\prime}\hat{\mathbf{B}}_{az}^{(2)}$ $e_{4k}^{(2)} = 0$ For $k \notin r_4^{(2)}$: $d_{4k}^{(2)} = 0$, $e_{4k}^{(2)} = 0$	For $k \in r_4^{(3)}$: $d_{4k}^{(3)} = \mathbf{y}_k^{(3)\prime}\hat{\mathbf{B}}_z^{(3)} - \mathbf{x}_k^{(3)\prime}\hat{\mathbf{B}}_{az}^{(3)}$ $e_{4k}^{(3)} = 0$ For $k \notin r_4^{(3)}$: $d_{4k}^{(3)} = 0$, $e_{4k}^{(3)} = 0$
5	For $k \in r_5^{(1)} (= s_5)$: $d_{5k}^{(1)} = 0$, $e_{5k}^{(1)} = 0$	For $k \in r_5^{(2)} (= s_5)$: $d_{5k}^{(2)} = 0$, $e_{5k}^{(2)} = 0$	For $k \in r_5^{(3)}$: $d_{5k}^{(3)} = \mathbf{y}_k^{(3)\prime}\hat{\mathbf{B}}_z^{(3)} - \mathbf{x}_k^{(3)\prime}\hat{\mathbf{B}}_{az}^{(3)}$ $e_{5k}^{(3)} = 0$ For $k \notin r_5^{(3)}$: $d_{5k}^{(3)} = 0$, $e_{5k}^{(3)} = 0$

Appendix 3 Unbiasedness of (4.3)

As previously mentioned, conditional on s_v, $\mathbf{m}_v^{(1)},\ldots,\ \mathbf{m}_v^{(i)}$, the set $r_{vg}^{(i)}$ is an SI-sample of size $m_{vg}^{(i)}$ from the $n_{vg}^{(i)}$ elements in $s_{vg}^{(i)}$, $g=1,\ldots,G_v^{(i)}$, $i=1,\ldots,I$ and $v=1,\ldots,V^*$, under assumptions *A1* and *A2*,. Thus, under *A1* and *A2*, it follows that for $k\in s_{vg}^{(i)}$

$$E_{RD}[(R_{vk}^{(i)})^2\,|\,\mathbf{m}_v^{(i)},\ldots,\mathbf{m}_v^{(1)},s_v)=E_{RD}(R_{vk}^{(i)}\,|\,\mathbf{m}_v^{(i)},\ldots,\mathbf{m}_v^{(1)},s_v)=\hat{\theta}_{vg}^{(i)}=m_{vg}^{(i)}/n_{vg}^{(i)}=\hat{\theta}_{vkk}^{(i,i)}$$

Moreover, for $k\neq l$, $\{k,l\}\in s_{vg}^{(i)}$ it follows that

$$E_{RD}(R_{vk}^{(i)}R_{vl}^{(i)}\,|\,\mathbf{m}_v^{(i)},\ldots,\mathbf{m}_v^{(1)},s_v)=\hat{\theta}_{vg}^{(i)}(m_{vg}^{(i)}-1)/(n_{vg}^{(i)}-1)=\hat{\theta}_{vkl}^{(i,i)}$$

Thus, for $i=i'$ it follows that

$$E_{RD}\Big(\sum\sum_{s_v}\frac{R_{vk}^{(i)}R_{vl}^{(i)}u_{vk,U}^{(i)}u_{vl,U}^{(i)}}{\pi_{vkl}\hat{\theta}_{vkl}^{(i,i')}}\,|\,\mathbf{m}_v^{(i)},\ldots,\mathbf{m}_v^{(1)},s_v\Big)=\sum\sum_{s_v}\frac{u_{vk,U}^{(i)}u_{vl,U}^{(i')}}{\pi_{vkl}}$$

Furthermore, it follows that for any $\{k,l\}\in s_v$

$$E_{RD}(R_{vk}^{(i)}R_{vl}^{(i')}\,|\,\mathbf{m}_v^{(i')},\ldots,\mathbf{m}_v^{(1)},s_v)=R_{vk}^{(i)}E_{RD}(R_{vl}^{(i')}\,|\,\mathbf{m}_v^{(i')},\ldots,\mathbf{m}_v^{(1)},s_v)=R_{vk}^{(i)}\hat{\theta}_{vl}^{(i')}$$

for $i<i'$. Thus, it follows that

$$\begin{aligned}&E_{RD}\Big(\sum\sum_{s_v}\frac{R_{vk}^{(i)}R_{vl}^{(i')}u_{vk,U}^{(i)}u_{vl,U}^{(i')}}{\pi_{vkl}\hat{\theta}_{vkl}^{(i,i')}}\,|\,\mathbf{m}_v^{(i')},\ldots,\mathbf{m}_v^{(1)},s_v\Big)=\\&\sum\sum_{s_v}\frac{R_{vk}^{(i)}E_{RD}(R_{vl}^{(i')}\,|\,\mathbf{m}_v^{(i')},\ldots,\mathbf{m}_v^{(1)},s_v)u_{vk,U}^{(i)}u_{vl,U}^{(i')}}{\pi_{vkl}\hat{\theta}_{vk}^{(i)}\hat{\theta}_{vl}^{(i')}}=\\&\sum\sum_{s_v}\frac{R_{vk}^{(i)}u_{vk,U}^{(i)}u_{vl,U}^{(i')}}{\pi_{vkl}\hat{\theta}_{vk}^{(i)}}\end{aligned}$$

for $1\leq i<i'\leq I$. Repeating the step above, conditioning on $\mathbf{m}_v^{(i)},\ldots,\mathbf{m}_v^{(1)},s_v$, it follows that

$$
\begin{aligned}
&E_{RD}(\sum\sum\nolimits_{s_v}\frac{R_{vk}^{(i)}u_{vk,U}^{(i)}u_{vl,U}^{(i')}}{\pi_{vkl}\hat{\theta}_{vk}^{(i)}}\,|\,\mathbf{m}_v^{(i)},...,\mathbf{m}_v^{(1)},s_v)=\\
&\sum\sum\nolimits_{s_v}\frac{E_{RD}(R_{vk}^{(i)}\,|\,\mathbf{m}_v^{(i)},...,\mathbf{m}_v^{(1)},s_v)u_{vk,U}^{(i)}u_{vl,U}^{(i')}}{\pi_{vkl}\hat{\theta}_{vk}^{(i)}}=\\
&\sum\sum\nolimits_{s_v}\frac{u_{vk,U}^{(i)}u_{vl,U}^{(i')}}{\pi_{vkl}}
\end{aligned}
$$

Since

$$
E_{p_v}(\sum\sum\nolimits_{s_v}\frac{u_{vk,U}^{(i)}u_{vl,U}^{(i')}}{\pi_{vkl}})=\sum\sum\nolimits_U u_{vk,U}^{(i)}u_{vl,U}^{(i')}=(\sum\nolimits_U u_{vk,U}^{(i)})(\sum\nolimits_U u_{vl,U}^{(i')})
$$

for any combination of $i=1,...,I$ and $i'=1,...,I$, it thus follows that under assumptions *A1* and *A2*,

$$
E_{p_v}[E_{RD}(\sum_{i=1}^{I}\sum_{i'=1}^{I}\sum\sum\nolimits_{s_v}\frac{R_{vk}^{(i)}R_{vl}^{(i')}u_{vk,U}^{(i)}u_{vl,U}^{(i')}}{\pi_{vkl}\hat{\theta}_{vkl}^{(i,i')}}\,|\,s_v)]=(\sum_{i=1}^{I}\sum\nolimits_U u_{vk,U}^{(i)})^2
$$

Appendix 4 Results from a small-scale simulation study

In order to evaluate the properties of $\hat{V}(\hat{T}_z)$, as defined in (4.6), a small-scale simulation study was performed for the case $I = 3$ and $V = 3$. For the study, an artificial population U, of size $N = 10000$ was generated using the approach discussed by Axelson (2000). Thus, the population was generated using a slightly modified version of the method suggested by Vale and Maurelli (1983), which allows for the generation of data according to a multivariate nonnormal distribution with specified correlation structure and given marginal means, variances, and coefficients of skewness and kurtosis. In generating the population, the correlation structure and the univariate moments for the variables in the real-world population MU281 (Särndal, Swensson, Wretman, 1992, Appendix B, pp. 652-659) was used as input. Hence, although artificial, the population used in the simulation is similar to MU281 in terms of the univariate moments of the variables and the pairwise correlation structure. In what follows, the variables in the simulation population are referred to using the names of the corresponding variables in MU281.

For element $k \in U$, $\mathbf{y}_k^{(i)} = (1, \varepsilon_y^{(i)}\text{REV84}_k^{(i-1)})'$, $i = 1,2,3$, where $\text{REV84}_k^{(0)}$ refers to the value obtained for REV84 in the generation of U and $\varepsilon_y^{(i)} \sim N(\mu_{y^{(i)}}, \sigma^2_{y^{(i)}})$, with

$$N(\mu_{y^{(i)}}, \sigma^2_{y^{(i)}}) = \begin{cases} N(1,0) & i = 1 \\ N(1.1, 0.1) & i = 2 \\ N(1.3, 0.1) & i = 3 \end{cases}$$

If used, the $\mathbf{x}$-vector for element $k \in U$ at time $i = 1,2,3$, is given by $\mathbf{x}_k^{(i)} = (1, \varepsilon_x^{(i)}\text{S82}_k^{(i-1)})'$, where $\text{S82}_k^{(0)}$ refers to the value obtained for S82 in the generation of U and $\varepsilon_x^{(i)} \sim N(\mu_{x^{(i)}}, \sigma^2_{x^{(i)}})$, with

$$N(\mu_{x^{(i)}}, \sigma^2_{x^{(i)}}) = \begin{cases} N(1,0) & i = 1 \\ N(1, 0.05) & i = 2,3 \end{cases}$$

The study variable is given by $z_k^{(i)} = \text{RMT85}_k^{(0)}$, $i = 1,2,3$, where $\text{RMT85}_k^{(0)}$ refers to the value obtained in the generation of U. Generated once, the values of all variables remain fixed throughout the whole simulation exercise.

In order to evaluate $\hat{V}(\hat{T}_z)$ under both simple and more complex conditions, the more complex ones possibly being more relevant from a practical point of view, 16 different scenarios were studied. The scenarios are given by all possible combinations of no- and yes answers to the following four questions:

- Q1: Is the auxiliary vector $\mathbf{x}$ used in the estimation?
  - If no, all terms of $\hat{T}_z$ and $\hat{V}(\hat{T}_z)$ involving $\mathbf{x}$ and are dropped.
  - If yes, the $\mathbf{x}$-vector defined above is used.
- Q2: The first time data are to be collected for the elements in sample $k \in s_v$, do the elements have differentiated response behavior?

  - If no, all elements in s_v will respond with the same probability, $\theta_{vk}^{(i_v)} = \theta_v^{(i_v)} = 0.75$ for $k \in s_v$, where i_v denotes the first time point when data are to be collected for $k \in s_v$.
  - If yes, the population is divided into two groups after increasing size of the variable P75, with the third quartile of P75 used as threshold. Elements belonging to the same group respond with the same probability, i.e. $\theta_{vk}^{(i_v)} = \theta_{vg}^{(i_v)}$ for all $k \in s_{vg}$, $g = 1,2$, with

$$\theta_{vg}^{(i_v)} = \begin{cases} 0.8 & g = 1 \\ 0.6 & g = 2 \end{cases}$$

  - $\theta_{vk}^{(i_v)} = \theta_{vg}^{(i_v)}$ for all $k \in s_{vg}$, $g = 1,2$.
- Q3: If data are to be collected for $k \in s_v$ at time $i > i_v$, do the response behavior at time $i > i_v$ depend on the response behavior at time $i-1$?
  - If no, $\theta_{vk}^{(i)} = \theta_{vk}^{(i_v)}$ for $k \in s_v$ at time $i > i_v$.
  - If yes,

$$\theta_{vk}^{(i)} = \begin{cases} 0.9 & \text{for } k \in r_v^{(i-1)} \\ 0.25 & \text{for } k \notin r_v^{(i-1)} \end{cases}$$

    for $k \in s_v$ at time $i > i_v$
- Q4: Is stratified simple random sampling used?
  - If no, s_v is drawn from U using simple random sampling, with $N_v = 10000$ and $n_v = 200$, $v = 1,...,5$.
  - If yes, s_v is drawn from U using stratified simple random sampling, with two strata, defined after increasing size of CS82, with $N_{v,1} = 7500$, $N_{v,2} = N - N_{v,1}$, and $n_{v,h} = 100$, $h = 1,2$, for $v = 1,...,5$.

A total of $M = 5000$ simulation runs were carried out for each of the 16 scenarios. Under each scenario, let $\hat{T}_{z,m}$ and $\hat{V}(\hat{T}_{z,m})$ denote the estimates computed according to (3.13) and (4.6), respectively, for simulation $m = 1,..., M$, using response homogeneity groups that coincide with the true response distribution as implied by the combined answers to questions Q3 and Q4 above. Let $RB_{MC}[\hat{V}(\hat{T}_z)]$ denote the Monte Carlo relative bias of $\hat{V}(\hat{T}_z)$, defined as

$$RB_{MC}[\hat{V}(\hat{T}_z)] = 100[\frac{\sum_{m=1}^{M} \hat{V}(\hat{T}_{z,m}) / M}{S_{\hat{T}_z}^2} - 1]$$

where

$$S_{\hat{T}_z}^2 = \frac{\sum_{m=1}^{M} \hat{T}_{z,m}^2 - (\hat{T}_{z,m})^2 / M}{M-1}$$

and let $EC_{MC}[\hat{T}_z, \hat{V}(\hat{T}_z)]$ denote the Monte Carlo empirical coverage rate of an approximate 95 % confidence interval with endpoints given by $\hat{T}_z \pm 2[\hat{V}(\hat{T}_z)]^{0.5}$, i.e.

$$EC_{MC}[\hat{T}_z, \hat{V}(\hat{T}_z)] = 100 \frac{\sum_{m=1}^{M} I(CI_m \ni T_z)}{M}$$

where

$$I(CI_m \ni T_z) = \begin{cases} 1 & \text{if } \dfrac{(\hat{T}_{z,m} - T_z)^2}{\hat{V}(\hat{T}_{z,m})} \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

As each scenario is evaluated under circumstances such that the results in sections 3 and 4 are valid, $\hat{V}(\hat{T}_z)$ is approximately unbiased according to theory. This expectation is not gravely contradicted by the results on $RB_{MC}[\hat{V}(\hat{T}_z)]$, presented in table 1 below. Furthermore, the results on $EC_{MC}[\hat{T}_z, \hat{V}(\hat{T}_z)]$ indicate that the expected coverage rate of a confidence interval with endpoints given by $\hat{T}_z \pm 2[\hat{V}(\hat{T}_z)]^{0.5}$ is close to 95 % under each scenario.

Table 1
Simulation results

Scenario	Answer to				$RB_{MC}[\hat{V}(\hat{T}_z)]$	$EC_{MC}[\hat{T}_z, \hat{V}(\hat{T}_z)]$	No of runs when assumption *A2* was not fulfilled
	Q1	Q2	Q3	Q4	(%)	(%)	
1	N	N	N	N	0.9	95.0	0
2	N	N	N	Y	2.4	95.5	0
3	N	N	Y	N	0.1	95.2	0
4	N	N	Y	Y	2.9	95.7	0
5	N	Y	N	N	-5.1	94.6	1
6	N	Y	N	Y	-1.5	95.2	1
7	N	Y	Y	N	-0.5	95.0	0
8	N	Y	Y	Y	2.9	95.6	0
9	Y	N	N	N	-2.4	94.2	0
10	Y	N	N	Y	-0.9	95.0	1
11	Y	N	Y	N	0.0	95.5	0
12	Y	N	Y	Y	-1.6	94.9	0
13	Y	Y	N	N	-7.5	93.8	0
14	Y	Y	N	Y	-1.5	95.3	1
15	Y	Y	Y	N	-3.6	94.9	0
16	Y	Y	Y	Y	-2.3	94.8	0

Appendix 5 Overview of the LFS subsample-structure

Principle η	No. v	Subsample $s_{\eta,v}$ from annual sample for year	Data to be collected for $k \in s_{\eta,v}$, month i			
			1	4	7	10
1	1	$\tau-2$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$, $\mathbf{Y}_k^{(1)}$, $Z_k^{(1)}$			
1	4	$\tau-2$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$, $\mathbf{Y}_k^{(4)}$, $Z_k^{(4)}$		
1	7	$\tau-2$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$	$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$, $\mathbf{Y}_k^{(7)}$, $Z_k^{(7)}$	
1	10	$\tau-1$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$	$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$, $\mathbf{Y}_k^{(10)}$, $Z_k^{(10)}$
1	13	$\tau-1$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$	$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$
1	16	$\tau-1$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$	$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$
1	19	$\tau-1$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$	$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$
1	22	τ	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$	$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$
1	25	τ		$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$	$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$
1	28	τ			$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$
1	31	τ				$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$
2	1	$\tau-2$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$,			
2	4	$\tau-2$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$		
2	7	$\tau-2$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$	$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	
2	10	$\tau-1$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$	$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$
2	13	$\tau-1$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$	$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$
2	16	$\tau-1$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$	$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$
2	19	$\tau-1$	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$	$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$
2	22	τ	$\mathbf{x}_k^{(1)}$, $\mathbf{y}_k^{(1)}$	$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$	$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$
2	25	τ		$\mathbf{x}_k^{(4)}$, $\mathbf{y}_k^{(4)}$	$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$
2	28	τ			$\mathbf{x}_k^{(7)}$, $\mathbf{y}_k^{(7)}$	$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$
2	31	τ				$\mathbf{x}_k^{(10)}$, $\mathbf{y}_k^{(10)}$

Principle η	No. ν	Subsample $s_{\eta,\nu}$ from annual sample for year	Data to be collected for $k \in s_{\eta,\nu}$, month i			
			2	5	8	11
1	2	$\tau-2$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$, $\mathbf{Y}_k^{(2)}$, $Z_k^{(2)}$			
1	5	$\tau-2$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$, $\mathbf{Y}_k^{(5)}$, $Z_k^{(5)}$		
1	8	$\tau-2$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$	$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$, $\mathbf{Y}_k^{(8)}$, $Z_k^{(8)}$	
1	11	$\tau-1$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$	$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$, $\mathbf{Y}_k^{(11)}$, $Z_k^{(11)}$
1	14	$\tau-1$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$	$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$
1	17	$\tau-1$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$	$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$
1	20	$\tau-1$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$	$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$
1	23	τ	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$	$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$
1	26	τ		$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$	$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$
1	29	τ			$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$
1	32	τ				$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$
2	2	$\tau-2$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$			
2	5	$\tau-2$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$		
2	8	$\tau-2$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$	$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	
2	11	$\tau-1$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$	$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$
2	14	$\tau-1$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$	$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$
2	17	$\tau-1$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$	$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$
2	20	$\tau-1$	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$	$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$
2	23	τ	$\mathbf{x}_k^{(2)}$, $\mathbf{y}_k^{(2)}$	$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$	$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$
2	26	τ		$\mathbf{x}_k^{(5)}$, $\mathbf{y}_k^{(5)}$	$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$
2	29	τ			$\mathbf{x}_k^{(8)}$, $\mathbf{y}_k^{(8)}$	$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$
2	32	τ				$\mathbf{x}_k^{(11)}$, $\mathbf{y}_k^{(11)}$

Principle η	No. v	Subsample $s_{\eta,v}$ from annual sample for year	Data to be collected for $k \in s_{\eta,v}$, month i			
			3	6	9	12
1	3	$\tau-2$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$, $\mathbf{Y}_k^{(3)}$, $Z_k^{(3)}$			
1	6	$\tau-2$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$	$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$, $\mathbf{Y}_k^{(6)}$, $Z_k^{(6)}$		
1	9	$\tau-2$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$	$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$	$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$, $\mathbf{Y}_k^{(9)}$, $Z_k^{(9)}$	
1	12	$\tau-1$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$	$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$	$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$, $\mathbf{Y}_k^{(12)}$, $Z_k^{(12)}$
1	15	$\tau-1$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$	$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$	$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$
1	18	$\tau-1$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$	$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$	$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$
1	21	$\tau-1$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$	$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$	$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$
1	24	τ	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$	$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$	$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$
1	27	τ		$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$	$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$
1	30	τ			$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$
1	33	τ				$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$
2	3	$\tau-2$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$			
2	6	$\tau-2$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$	$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$		
2	9	$\tau-2$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$	$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$	$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	
2	12	$\tau-1$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$	$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$	$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$
2	15	$\tau-1$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$	$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$	$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$
2	18	$\tau-1$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$	$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$	$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$
2	21	$\tau-1$	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$	$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$	$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$
2	24	τ	$\mathbf{x}_k^{(3)}$, $\mathbf{y}_k^{(3)}$	$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$	$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$
2	27	τ		$\mathbf{x}_k^{(6)}$, $\mathbf{y}_k^{(6)}$	$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$
2	30	τ			$\mathbf{x}_k^{(9)}$, $\mathbf{y}_k^{(9)}$	$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$
2	33	τ				$\mathbf{x}_k^{(12)}$, $\mathbf{y}_k^{(12)}$

2006:1 Sysselsättningsavgränsning i RAMS – Metodöversyn 2005

2007:1 Några väsentliga sysselsättningsbegrepp i den officiella statistiken

2007:2 Registerbaserad aktivitetsstatistik

2008:1 Ungdomar utan fullföljd gymnasieutbildning – en undersökning med många utmaningar

2009:1 Longitudinell Integrationsdatabas för Sjukförsäkrings- och Arbetsmarknadsstudier (LISA) 1990–2007

2010:1 Principiella grunder för Arbetskraftsundersökningarna (AKU) och arbetsmarknadsstatistiken

2010:2 Rekryteringsstatistik från AKU

2011:1 Basic principles for Labour Force Surveys (LFS) and labour market statistics

2011:2 Recruitment Statistics for the Swedish Labour Force Survey

2011:3 Arbetskraftsundersökningarna (AKU) 50 år. Fyra forskarperspektiv på arbetsmarknaden

2011:4 Longitudinell integrationsdatabas för Sjukförsäkrings- och Arbetsmarknadsstudier (LISA) 1990–2009

2011:5 Yrkesregistret med yrkesstatistik. En beskrivning av innehåll och kvalitet

2011:6 Urvals- och estimationsförfarandet i de svenska arbetskraftsundersökningarna (AKU)

2011:7 Konjunkturberoende i inflödet till och utflödet från högre studier

2012:1 Actual Hours Worked in the Swedish LFS. Four articles

2013:1 Ungdomsarbetslöshet – jämförbarhet i statistiken mellan ett antal europeiska länder

2013:2 Revising Surveys – Linking Old and New Data

2013:3 Consistent Seasonal Adjustment and Trend-cycle Estimation

2013:4 Youth unemployment – comparability in statistics between a number of European countries

2013:5 Länkning av centrala serier i Arbetskraftsundersökningarna (AKU) 1970–1986

2013:6 Deltidssysselsatta arbetssökande

2013:7 Part-time employed job seekers

2013:8 Arbetskraftsundersökningarnas (AKU) och Arbetsförmedlingens (Af) arbetslöshetsstatistik – En jämförande studie

2014:1 Metod för estimation vid sammanslagning av urval med olika design i arbetskraftsundersökningarna

2014:2 Mätfelsstudie i AKU

2015:1 Flow statistics – an early indicator of turning points on the labour market?

ISSN 1654-465X (Online)
ISSN 1103-7458 (Print)
ISBN 978-91-618-1628-6 (Print)