SAPIENZA
UNIVERSITÀ DI ROMA

Ph.D. in Methodological Statistics

Doctoral Thesis

Department of Statistical Sciences

Estimation of the variance for different estimators of the change over time for overlapping samples

Candidate:
Diego Chianella

Thesis advisor:
Prof. Pier Luigi Conti

Roma – June 2019

Ringraziamenti

Vorrei ringraziare tutte le persone che mi hanno accompagnato in questo percorso.

Un primo grande ringraziamento va a Fabio Bacchini e Roberto Iannaccone che mi hanno incoraggiato nel mio lavoro di ricerca e dato la possibilità di crescere grazie a preziosi confronti e consigli, motivandomi ad intraprendere il percorso di Dottorato.

Un sentito ringraziamento a Giancarlo Bruno e al servizio Istat sulle statistiche congiunturali sulle imprese per avermi permesso l'elaborazione dei dati utili al mio lavoro di ricerca, così come a Claudio Ceccarelli e al servizio Istat sul sistema integrato lavoro, istruzione e formazione per le diverse ed interessanti occasioni di ricerca che hanno favorito la mia crescita in ambito metodologico.

Ringrazio anche i numerosi colleghi e colleghe dell'Istat, fonte di continui stimoli. Un ringraziamento speciale va ad Agostina Zanoli, Barbara Guardabascio, Cristina Ciaffarafà, Giulia Ippoliti e Marica D'Elia alle quali mi lega anche un rapporto di sincera amicizia.

Un grazie a tutti i ricercatori, i docenti e i dottorandi del Dipartimento di Scienze Statistiche con cui ho avuto la possibilità di relazionarmi in questi anni e al Coordinatore del Dottorato Pier Luigi Conti, per aver seguito e supervisionato il mio lavoro di tesi.

Grazie anche ad Alberto Sabbi, collega, amico e premuroso compagno di studi con cui ho condiviso gioie e dolori di questa esperienza.

Grazie alla mia famiglia e a tutti i miei amici per essere sempre presenti.

Infine un ringraziamento speciale è rivolto ad Ilaria per essermi stata vicino e avermi motivato durante i lunghi mesi di lavoro.

*Nessuna umana investigazione si
può dimandare vera scienza, se
essa non passa per le matematiche
dimostrazioni e se tu dirai che le
scienze, che principiano e finiscono
nella mente, abbiano verità, questo
non si concede, ma si nega per molte
ragioni; e prima, che in tali discorsi
mentali non accade esperienza, senza
la quale nulla dà di sé certezza.*

Leonardo Da Vinci

Contents

CHAPTER 3

CHAPTER 4

Introduction

This work was inspired by the growing need to have a measure of the accuracy of the estimates produced within the short-term statistics in the Official Statistics. In particular, the aim of the work is to illustrate the methodology for the computation of the variance for the estimators currently used in the service turnover survey carried on by the Italian National Institute of Statistics (ISTAT) for the quarterly turnover growth rate estimation. The variance for the estimators currently used in the service turnover survey is computed only for the total estimations in the quarters t and t-4, while the variance of the growth rate estimation for the different estimation domains is not calculated. My methodological contribution is not only to suggest how to assess the variance of possible estimators of the turnover variation over time, but also to compare such estimators with respect to their variance to identify the best one.

While the adopted methodologies are fairly uniform within structural statistics on companies, this does not happen for short-term statistics, where the situation is quite heterogeneous. In fact, at European level, as indicated in the Short-Term Methodologies Handbook by Eurostat (see Eurostat, 2006), the choice of methodologies to be implemented is left to the various National Statistical Institutes. This heterogeneity appears both at the sampling plan level and at the estimation methods level.

Short-term statistics measure the evolution of a phenomenon over time. Often, we are not interested in the value itself of the variables of interest, but rather in their variation over time. Changes can be measured as the difference between two quantities at different waves (for example the difference in unemployment rate between two consecutive quarters) or as the relative percentage difference between two quantities over time (e.g. the percentage change of turnover with respect to previous quarter or same quarter of the previous year). In these cases, the variance is important for the production of a confidence interval of the variation. Confidence intervals are useful not only to evaluate the reliability of the estimate, but also to understand if a variation is statistically significant. In fact, if the confidence interval does not contain the zero value, it means that the calculated variation is statistically different from zero.

While the calculation of the variance of the estimates produced for a given instant of time is now a good practice (also through the development of software packages), the same does not happen for the variation of two quantities over time. An estimator of variance must take into account of both the estimator and the sampling design (Wolter, K.M. (1985)). The biggest difficulty is that for many surveys, the samples for producing estimates in two different time are not independent each other, due to the rotation operations of the sample. In particular for business surveys, in order to take into account the birth-mortality of units in the population and changes in stratification variables (such as size category and type of economic activity), the sample is updated, and a part of the units is replaced with others. Surveys, such as the Italian EU-SILC survey and the Italian Labour Force survey (LFS), include a rotated panel sample, resulting in partially overlapping samples between two occasions (Gazzelloni (2006), Ceccarelli et al. (2008)). This means that in calculating the estimate of the variance of change over time, we need not only the estimates of the variances of the cross-sectional estimates, but also the covariance terms between cross-sectional estimates. Moreover, many indicators are non-linear functions of linear estimators (e.g. simple ratio, difference of ratios), therefore, to calculate their variance a first-order Taylor approximation can be used. This is the case, for example, for the variance estimations of the LFS-based indicators' annual net changes (Ceccarelli et al. (2017)). Alternatively, balanced repeated replication (BRR) can be used (Moretti et al. (2005)).

Currently, two estimators for the turnover growth rate estimation in the domain of services are used. The first is based on the variation computed on the overlapping sample units in both occasions (the quarter t and the quarter t-4), while the second is based on the ratio of totals computed through calibration, using all observations in both quarters and not only the overlapping sample units (Bacchini et al. (2013), Chianella et al. (2013), Bacchini et al. (2014), Bacchini et al. (2015), Chianella et al. (2015)). Other two estimators are taken into consideration in this study. The first additional estimator is based on the ratio of totals computed through calibration using only the overlapping sample units in both occasions, while the second additional estimator is based on the ratio of the sample means calculated using turnover data on all respondent units over the two quarters. Therefore, four non-linear estimators are presented for the turnover growth rate: ratio of sample means at the periods t and t-4, and ratio of totals through calibration, both computed: (i) on all the respondents units at both occasions, and (ii) only on overlapping respondents units. The performance is assessed by a simulation study, which also has the aim of exploring under which conditions it is better to use all the observations or only the overlapping observations. The change estimators and the corresponding estimators of the variance are defined at stratum and estimation domain level and take into account the use of a stratified sampling design and the updating of the sample due to a replacement of some units and to a dynamic stratification of the population.

This work is organized as follows.

The first Chapter provides an overview of the literature available about the variance of the change over time. Contributions of different authors are described, focusing the attention on different types of population (large/not large population; stratified/non-stratified population, with/without the hypothesis of no birth-mortality in the population).

Chapter 2 describes the methodology used in Istat for the quarterly turnover growth rate estimation in the service sector. The aim of this Chapter is to introduce the methodology for the computation of the standard errors for the quarterly turnover growth rate. The computation is performed using the Taylor series approximation, at stratum and estimation domain level. It is also provided the formula of the overlapping values over which the estimator using only the overlapping sample units between both occasions is better than the estimator using all observations in both occasions.

In Chapter 3, a simulation study was conducted with the aim of analyzing the performance of these estimators and exploring under which conditions it is better to use all the observations or only the overlapping observations. The bias, the standard deviation and the mean squared error have been analyzed through 1000 different samples extracted from the population, considering different values of the overlap between the respondent units at the occasions *t* and *t-4* and different values of the correlation between the variable of interest and the calibration variable, together with different correlations between Y^t and Y^{t-4}. The estimator with minimum mean squared error was preferred.

In Chapter 4 an application performed on real data is described, using information from the quarterly service turnover survey with the aim to evaluate the standard errors associated with different estimates. A confidence interval is defined at 95% level. The standard errors obtained with the Taylor series approximation are compared with those obtained with the bootstrap method. The results are also compared with the results obtained by Knottnerus and Van Delden (2012) about the standard error of the turnover growth rate in Dutch supermarkets.

CHAPTER 1

Literature on the variance of the change over time

1.1 - The variance of change based on overlapping samples from large populations

Suppose we are interested in estimating the change in the mean value in the population, of a quantity on two different occasions $d = \bar{Y}_2 - \bar{Y}_1$. We use the difference between sample means calculated on two different occasions:

$$\hat{d} = \bar{y}_2 - \bar{y}_1.$$

The variance of the difference, can be expressed as:

$$Var(\hat{d}) = Var(\bar{y}_2) + Var(\bar{y}_1) - 2\text{Cov}\ (\bar{y}_2, \bar{y}_1) =$$

$$= \ Var(\bar{y}_2) + Var(\bar{y}_1) - 2\text{Cov}\ \left(\frac{y_2}{n_2}, \frac{y_1}{n_1}\right) =$$

$$Var(\bar{y}_2) + Var(\bar{y}_1) - \frac{2}{n_2 n_1}\text{Cov}\ (y_2, y_1).$$

where $y_1 = \sum_{i=1}^{n_1} y_{1i}$ and $y_2 = \sum_{i=1}^{n_2} y_{2i}$.

Considering a simple random sample without replacement (srswor), with a fixed size of the sample n, it is shown (Kish 1965, p. 63) that the variance and the covariance terms, are equal to:

$$Var(\bar{y}) = \frac{(1-f)}{n}\ S_{y_2}^2$$

$$Cov\ (y_2, y_1) = \ (1-f)nS_{y_1y_2}\ ,$$

where f and n are the sample fraction and the sample size, respectively. In the general case, where we have two samples of different size, assuming that the population is the same over time (there is no birth-mortality), Kish (1965, pp. 457-466) obtains the general formula:

$$\widehat{Var}(\hat{d}) = \frac{(1-f_2)}{n_2}\hat{S}^2_{y_2} + \frac{(1-f_1)}{n_1}\hat{S}^2_{y_1} - 2(1-f)\frac{n_c}{n_1n_2}\hat{S}_{y_1y_2} =$$

$$= \left(\frac{1}{n_2}-\frac{1}{N}\right)\hat{S}^2_{y_2} + \left(\frac{1}{n_1}-\frac{1}{N}\right)\hat{S}^2_{y_1} - 2n_c\left(\frac{1}{n_c}-\frac{1}{N}\right)\frac{n_c}{n_1n_2}\ \hat{S}_{y_1y_2}\ ,$$

where $f = n_c/N$ and n_c is the size of the overlap units to both samples s_1, s_2 (Figure 1.1). In the case of large population, the above expression becomes :

$$\widehat{Var}(\hat{d}) = \frac{\hat{S}^2_{y_2}}{n_2} + \frac{\hat{S}^2_{y_1}}{n_1} - \frac{2}{n_2n_1}\ n_c\ \hat{R}_{y_1y_2}\hat{S}_{y_1}\hat{S}_{y_2} =$$

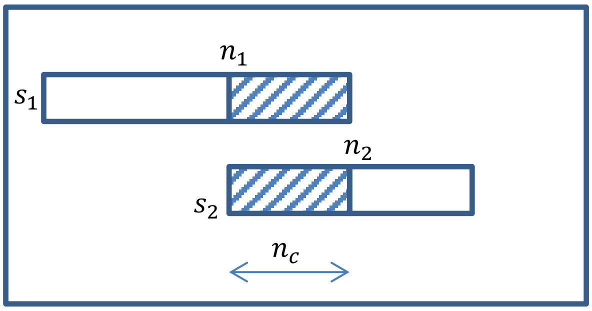
$$\frac{\hat{S}^2_{y_2}}{n_2} + \frac{\hat{S}^2_{y_1}}{n_1} - n_c\frac{n_c}{n_c}\frac{2\hat{R}_{y_1y_2}\hat{S}_{y_1}\hat{S}_{y_2}}{n_2n_1} =$$

$$= \frac{\hat{S}^2_{y_2}}{n_2} + \frac{\hat{S}^2_{y_1}}{n_1} - \frac{2o_1o_2\hat{R}_{y_1y_2}\hat{S}_{y_1}\hat{S}_{y_2}}{n_c},$$

where the covariance term is written as the product between the correlation coefficient estimated from the common sample $\hat{R}_{y_1y_2}$and the standard deviation $\hat{S}_{y_1}\ \hat{S}_{y_2}$, while the overlap in the first sample and in the second sample, are defined as:

$$o_1 = \frac{n_c}{n_1}\ ;\ o_2 = \frac{n_c}{n_2}\ .$$

Figure 1.1 - Two samples of different sizes ($\mathbf{n_1}$and $\mathbf{n_2}$) with overlap of size $\mathbf{n_c}$



Now, when $\hat{S}^2_{y_1} = \hat{S}^2_{y_2} = \hat{S}^2_y$ and $n_1 = n_2 = n$, the formula of the variance becomes:

$$\widehat{Var}(\hat{d}) = \frac{\hat{S}^2_y}{n} + \frac{\hat{S}^2_y}{n} - \frac{2n_c\hat{R}_{y_1y_2}\hat{S}_y\hat{S}_y}{n\ n} =$$

$$= \frac{2\hat{S}_y^2}{n} - \frac{2}{n}\, o\hat{R}_{y_1y_2}\hat{S}_y{}^2 =$$

$$\frac{2\hat{S}_y^2}{n}\left(1 - o\hat{R}_{y_1y_2}\hat{S}_y{}^2\right).$$

If the correlation $\hat{R}_{y_1y_2}$is positive, when we have complete overlap between the two sample at the different occasions ($o = \frac{n_c}{n} = 1$), then the variance of the difference $\hat{d}$ will take its smallest value.

From the general case, Kish reports four particular cases, where the first two, concern the extreme cases of non overlap and complete overlap between the samples:

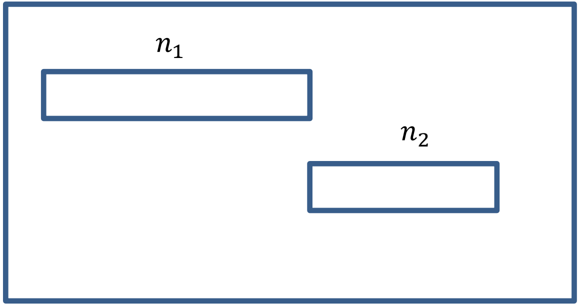
- **Case 1. There is no overlap between the two samples ($\boldsymbol{n_c} = \mathbf{0}$)**

In this case (Figure 1.2), we have that $\mathrm{o}_1 = \mathrm{o}_2 = 0$ and from the general formula of the variance of the difference, we obtain:

$$\widehat{Var}(\hat{d}) = \frac{\hat{S}_{y_2}^2}{n_2} + \frac{\hat{S}_{y_1}^2}{n_1}.$$

Therefore in this case the covariance term does not attend to lower $\widehat{Var}(\hat{d})$

Figure 1.2 - Non overlap between the sample

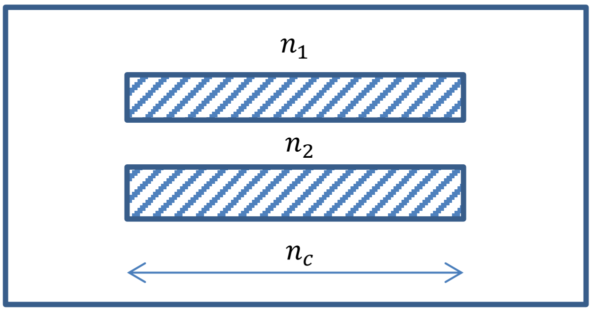


- **Case 2. There is complete overlap between the two samples ($\boldsymbol{n_c} = \boldsymbol{n_1} = \boldsymbol{n_2}$).**

In this case (Figure 1.3), we have that $o_1 = o_2 = 1$. The formula of the variance becomes:

$$\widehat{Var}(\hat{d}) = \frac{1}{n}\left(\hat{S}_{y_2}^2 + \hat{S}_{y_1}^2 - 2\hat{R}_{y_1y_2}\hat{S}_{y_1}\hat{S}_{y_2}\right).$$

Figure 1.3 - Complete overlap between the sample



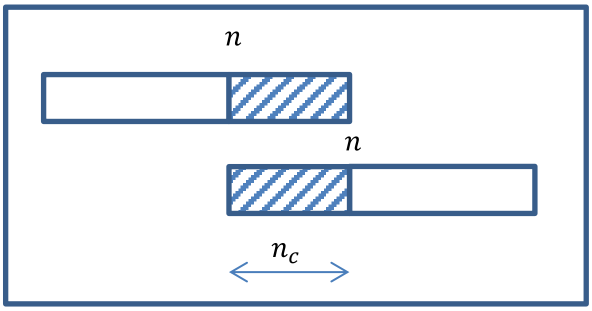
- **Case 3. The samples have identical size and partial overlap ($n_1 = n_2 = n > n_c$)**

In this case (Figure 1.4), $o_1 = o_2 = \text{o} < 1$. The formula for the variance of the difference can be written:

$$\widehat{Var}(\hat{d}) = \frac{\hat{S}_{y_2}^2}{n} + \frac{\hat{S}_{y_1}^2}{n} - \frac{2n_c\hat{R}_{y_1y_2}\hat{S}_{y_1}\hat{S}_{y_2}}{n^2} = \frac{1}{n}\left(\hat{S}_{y_2}^2 + \hat{S}_{y_1}^2 - 2\ o\ \hat{R}_{y_1y_2}\hat{S}_{y_1}\hat{S}_{y_2}\right).$$

Compared to case 2, we have now the term o in the formulation. Since $o < 1$, $\widehat{Var}(\hat{d})$ will be higher with respect to the case of complete overlap between the samples.

Figure 1.4 - Samples with the same size and partial overlap



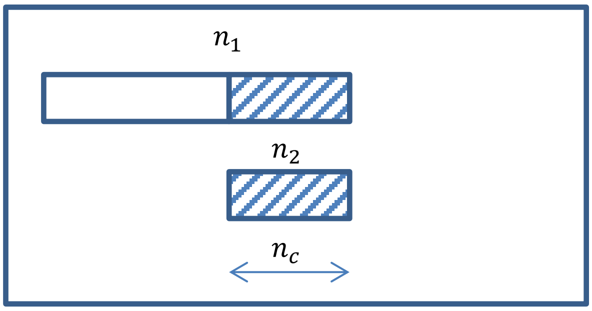
- **Case 4. The sample at the second occasion is a subset of the first**

In this case (Figure 1.5), we have $n_c = n_2 < n_1$. It follows that $o_2 = 1$ and $o_1 < 1$. The formula for the variance, becomes:

$$\widehat{Var}(\hat{d}) = \frac{\hat{S}_{y_2}^2}{n_c} + \frac{\hat{S}_{y_1}^2}{n_1} - \frac{2n_c\hat{R}_{y_1y_2}\hat{S}_{y_1}\hat{S}_{y_2}}{n_1n_c} = \frac{\hat{S}_{y_1}^2}{n_c} + \frac{\hat{S}_{y_1}^2 - 2\hat{R}_{y_1y_2}\hat{S}_{y_1}\hat{S}_{y_2}}{n_1}.$$

From this formulation can be noted that if $(\hat{S}_{y_1}^2 - 2\widehat{R}_{y_1 y_2}\hat{S}_{y_1}\hat{S}_{y_2}) < 0$, then more increase the size of the units not in overlapping (more n_1 is greater than n_c) and more increase the value of $\widehat{Var}(\hat{d})$.

Figure 1.5. The second sample is a subset of the first



1.2 - The variance of change without the hypothesis of large population

The computation of the variance of $\hat{d}$ becomes more complicated when the hypotesis of large population is removed. In fact, in finite populations, two disjoint samples are not independent.

Tam (1984) formulated the exact expression for the sampling variance of the difference $\hat{d}$, removing the hypothesis of "large" population supposed in Kish. The finite population corrections now are not negligible. The general formula of a srswor in this case is:

$$\widehat{Var}(\hat{d}) = \left(1 - \frac{n_1}{N}\right)\frac{\hat{S}_{y_1}^2}{n_1} + \left(1 - \frac{n_2}{N}\right)\frac{\hat{S}_{y_2}^2}{n_2} - 2(1-f)\frac{n_c}{n_1 n_2}\hat{S}_{y_1 y_2}.$$

From the general formula Tam derived the formulas for three different sampling plans, assigning different values of f.

1. In the sampling plan A, at the first occasion we have a srswor of size n_1 from a population U. In the second occasion we have a sample consisting in the union of a random subset of the first sample s_1 (s_c, of fixed size $n_c < n_1$), and a srswor from U excluding the units in the first sample s_1 (see Figure 1.6). In this case $\boldsymbol{f} = \frac{\boldsymbol{n_1 n_2}}{\boldsymbol{N n_c}}$ and:

$$\widehat{Var}(\hat{d}) = \left(1-\frac{n_1}{N}\right)\frac{\hat{S}^2_{y_1}}{n_1} + \left(1-\frac{n_2}{N}\right)\frac{\hat{S}^2_{y_2}}{n_2} - 2\left(1-\frac{n_1 n_2}{N n_c}\right)\frac{n_c}{n_1 n_2}\hat{S}_{\mathrm{y}_1\mathrm{y}_2} =$$

$$= \left(\frac{1}{n_1}-\frac{1}{N}\right)\hat{S}^2_{y_1} + \left(\frac{1}{n_2}-\frac{1}{N}\right)\hat{S}^2_{y_2} - 2\left(\frac{n_c}{n_1 n_2}-\frac{1}{N}\right)\hat{S}_{\mathrm{y}_1\mathrm{y}_2}$$

If the size of the samples is the same in both occasions ($n_1 = n_2 = n$), we can write:

$$\widehat{Var}(\hat{d}) = \left(\frac{1}{n}-\frac{1}{N}\right)\left(\hat{S}^2_{y_1} + \hat{S}^2_{y_2}\right) - 2\left(\frac{o}{n}-\frac{1}{N}\right)\hat{S}_{\mathrm{y}_1\mathrm{y}_2}$$

2. In the sampling plan B, at the first occasion we have a srswor of size n_1 from a population U. In the second occasion we have a sample consisting in the union of a random subset of the first sample s_1 (s_c, of fixed size $n_c < n_1$) and a srswor from U excluding the units in the overlap sample s_c (see Figure 1.7). In this case $\boldsymbol{f} = \frac{\boldsymbol{n_1 n_2}}{\boldsymbol{N n_c}} - \frac{(\boldsymbol{n_2 - n_c})(\boldsymbol{n_1 - n_c})}{(\boldsymbol{N - n_c})\boldsymbol{n_c}}$ and:

$$\widehat{Var}(\hat{d}) =$$

$$= \left(1-\frac{n_1}{N}\right)\frac{\hat{S}^2_{y_1}}{n_1} + \left(1-\frac{n_2}{N}\right)\frac{\hat{S}^2_{y_2}}{n_2} - 2\left(1-\frac{n_1 n_2}{N n_c} - \frac{(n_2-n_c)(n_1-n_c)}{(N-n_c)n_c}\right)\frac{n_c}{n_1 n_2}\hat{S}_{\mathrm{y}_1\mathrm{y}_2} =$$

$$= \left(\frac{1}{n_1}-\frac{1}{N}\right)\hat{S}^2_{y_1} + \left(\frac{1}{n_2}-\frac{1}{N}\right)\hat{S}^2_{y_2}$$

$$- 2\left(\frac{n_c}{n_1 n_2} - \frac{1}{N} - \frac{(n_2-n_c)(n_1-n_c)}{(N-n_c)\, n_1 n_2}\right)\hat{S}_{\mathrm{y}_1\mathrm{y}_2}$$

3. In the sampling plan C, at the first occasion we have a srswor of size n_1 from a population U. In the second occasion the first sample is replaced with a srswor from U (see Figure 1.8). In this case n_c is random and $\boldsymbol{f} = \mathbf{1}$ and:

$$\widehat{Var}(\hat{d}) = \left(1-\frac{n_1}{N}\right)\frac{\hat{S}^2_{y_1}}{n_1} + \left(1-\frac{n_2}{N}\right)\frac{\hat{S}^2_{y_2}}{n_2} =$$

$$= \left(\frac{1}{n_1}-\frac{1}{N}\right)\hat{S}^2_{y_1} + \left(\frac{1}{n_2}-\frac{1}{N}\right)\hat{S}^2_{y_2}$$

As we can see in the above formulations, the smallest value of the variance of the difference between two cross sectional estimates, when there is overlap between samples, is obtained in the sampling plan B. This sampling plan is similar to that in the Italian survey for turnover (we will discuss on it in next chapters), with the difference that at the second occasion the subset of the first sample is not completely random but there is a purposive choice.

Figure 1.6. Sampling plan A

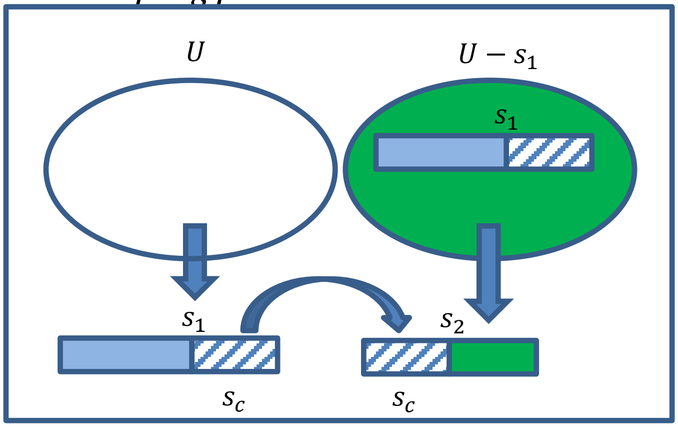


Figure 1.7. Sampling plan B

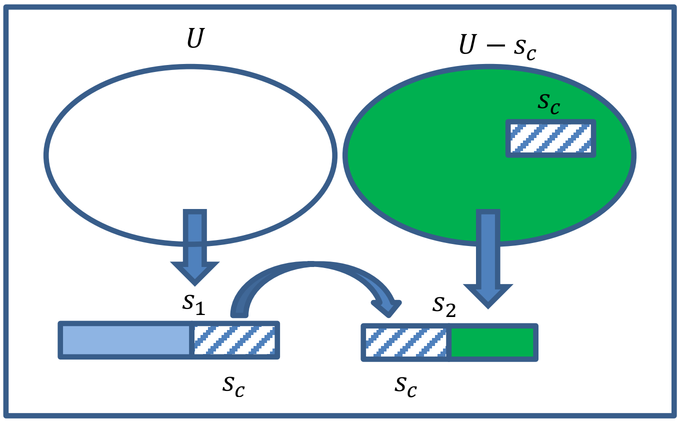
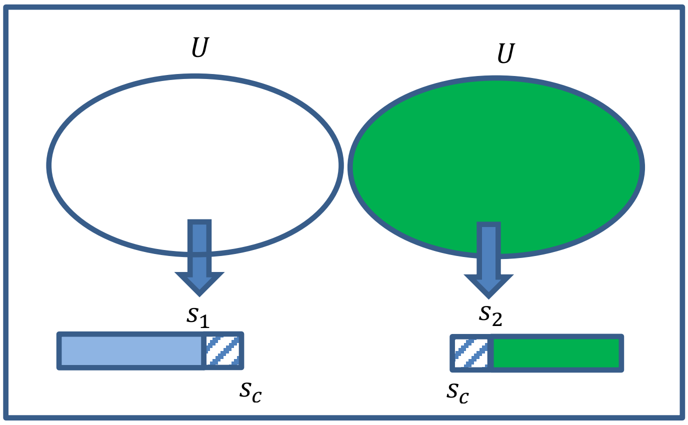
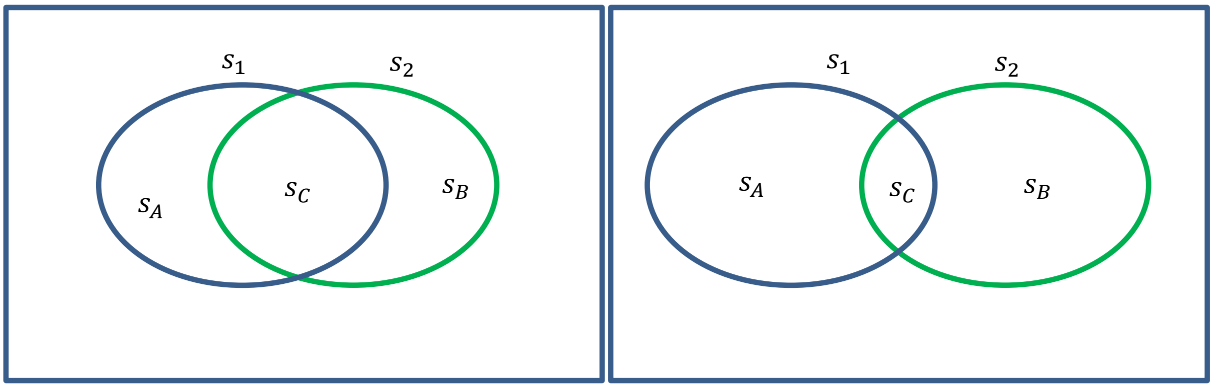


Figure 1.8. Sampling plan C



Qualitè and Tillé (2008), also took into account two samples, s$_1$ and s$_2$, selected without replacement and of fixed size n_1 and n_2 respectively. Then they considered $s_A = s_1|s_2$, $s_B = s_2|s_1$ and $s_C = s_1 \cap s_2$, of random size (see Figure 1.9). The Horvitz-Thompson estimator is used for calculate the totals $y_1$and y_2 and considered the difference of totals $\hat{\Delta}= y_2 - y_1$. To compare the results with Tam we instead consider the difference $\hat{d} = \bar{y}_2 - \bar{y}_1$.

Figure 1.9. Overlapping between two samples of fixed size of s_1, s_2



Then we have:

$$Var(\hat{d}) = Var(\bar{y}_2) + Var(\bar{y}_1) - 2Cov\ (\bar{y}_2, \bar{y}_1),$$

we can write:

$$Cov\ (\bar{y}_2, \bar{y}_1) = E[cov(\bar{y}_1, \bar{y}_2 | n_A, n_B, n_C)] + cov[E(\bar{y}_1 | n_A, n_B, n_C), \bar{y}_2 | n_A, n_B, n_C)].$$

Now, $\bar{y}_1$and $\bar{y}_2$ are unbiased conditional on n_A, n_B, n_C, thus:

$$cov[E(\bar{y}_1 | n_A, n_B, n_C), \bar{y}_2 | n_A, n_B, n_C)] = cov(\bar{Y}_1, \bar{Y}_2) = 0\,.$$

Furthermore, since we are in the sampling plan A of Tam, thus, as we show in the previous pages, the conditional covariance is:

$$cov(\bar{y}_1, \bar{y}_2 | n_A, n_B, n_C) = \left(\frac{n_c}{n_1 n_2} - \frac{1}{N}\right) \hat{S}_{y_1 y_2}\,,$$

And hence:

$$Cov\ (\bar{y}_2, \bar{y}_1) = E\left[\left(\frac{n_c}{n_1 n_2} - \frac{1}{N}\right) S_{y_1 y_2}\right] = \left(\frac{E(n_c)}{n_1 n_2} - \frac{1}{N}\right) S_{y_1 y_2}\,.$$

In this way the estimation of the variance of d̂ becomes:

$$\widehat{Var}(\hat{d}) = \left(\frac{1}{n_1} - \frac{1}{N}\right)\hat{S}^2_{y_1} + \left(\frac{1}{n_2} - \frac{1}{N}\right)\hat{S}^2_{y_2} - 2\left(\frac{E(n_c)}{n_1 n_2} - \frac{1}{N}\right)\hat{S}_{y_1 y_2}\,.$$

Where $\hat{S}_{y_1y_2}$is the simple covariance, calculated from the sample s_C:

$$\hat{S}_{y_1y_2} = \frac{1}{n_c - 1}\sum_{s_c}(y_{1k} - \bar{y}_{1c})\,(y_{2k} - \bar{y}_{2c})\,.$$

A few comment on the above formula are in order.

1. If the two samples s_1, s_2 form a panel we have that $n_1 = n_2 = n_c = n$ and we obtain:

$$\widehat{Var}(\hat{d}) = \left(\frac{1}{n} - \frac{1}{N}\right)\hat{S}^2_{y_1} + \left(\frac{1}{n} - \frac{1}{N}\right)\hat{S}^2_{y_2} - 2\left(\frac{E(n)}{n * n} - \frac{1}{N}\right)\hat{S}_{y_1y_2} =$$

$$\left(\frac{1}{n} - \frac{1}{N}\right)\left(\hat{S}^2_{y_1} + \hat{S}^2_{y_2} - 2\hat{S}_{y_1y_2}\right).$$

2. If the two samples s_1, s_2, are disjoint $(n_c = 0)$ then $E(n_c) = 0$ and:

$$\widehat{Var}(\hat{d}) = \left(\frac{1}{n} - \frac{1}{N}\right)\hat{S}^2_{y_1} + \left(\frac{1}{n} - \frac{1}{N}\right)\hat{S}^2_{y_2} + \frac{2}{N}\hat{S}_{y_1y_2}$$

In this case, unlike the case 1 on Kish, conditionally on $n_c = 0$ can lead the covariance term to increase the variance of the change $\hat{d}$ (if the correlation between $y_1$and y_2 is positive).

Qualitè and Tillé also compare the variance of the estimator $\hat{d}$ with the estimator $\hat{d}_c = \bar{y}_{2c} - \bar{y}_{1c}$, that considers the difference between the two cross sectional estimators only on the overlap between the samples. Also $\hat{d}_c$ is unbiased conditional on n_C.

Using the approximation $E\left(\frac{1}{n_c}\right) \approx \frac{1}{E(n_c)}$, the unconditional variance of $\hat{d}_c$ is:

$$\widehat{Var}(\hat{d}_c) = \left[E\left(\frac{1}{n_c}\right) - \frac{1}{N}\right]\left(\hat{S}^2_{y_1} + \hat{S}^2_{y_2}\right) - 2\left[\frac{E(n_c)}{E(n_c)E(n_c)} - \frac{1}{N}\right]\hat{S}_{y_1y_2} =$$
$$= \left[E\left(\frac{1}{n_c}\right) - \frac{1}{N}\right]\left(\hat{S}^2_{y_1} + \hat{S}^2_{y_2} - 2\hat{S}_{y_1y_2}\right).$$

The difference $Var(\hat{d}) - Var(\hat{d}_c)$ is:

$$Var(\hat{d}) - Var(\hat{d}_c) = \left(\frac{1}{n_1} - \frac{1}{N}\right)\hat{S}^2_{y_1} + \left(\frac{1}{n_2} - \frac{1}{N}\right)\hat{S}^2_{y_2} - 2\left(\frac{E(n_c)}{n_1n_2} - \frac{1}{N}\right)\hat{S}_{y_1y_2}$$

$$-\left(E\left(\frac{1}{n_c}\right)-\frac{1}{N}\right)\hat{S}^2_{y_1}-\left(E\left(\frac{1}{n_c}\right)-\frac{1}{N}\right)\hat{S}^2_{y_2}+2\left(E\left(\frac{1}{n_c}\right)-\frac{1}{N}\right)\hat{S}_{y_1y_2}=$$

$$=\left(\frac{1}{n_1}-E\left(\frac{1}{n_c}\right)\right)\hat{S}^2_{y_1}+\left(\frac{1}{n_2}-E\left(\frac{1}{n_c}\right)\right)\hat{S}^2_{y_2}-2\left(\frac{E(n_c)}{n_1n_2}-E\left(\frac{1}{n_c}\right)\right)\hat{S}_{y_1y_2}.$$

In the case $n_1 = n_2 = n$, $\hat{S}^2_{y_1} = \hat{S}^2_{y_2} = S^2$ the authors obtained:

$$Var(\hat{d})-Var(\hat{d}_c)\approx\left(\frac{1}{n}-\frac{1}{E(n_c)}\right)2\hat{S}^2-\left(\frac{E(n_c)}{n*n}-\frac{1}{E(n_c)}\right)2\rho\hat{S}^2.$$

Recalling that the overlap is $o=\frac{n_c}{n}$ (see the case 3 of Kish), in this case we can write:

$$o=\frac{E(n_c)}{n},$$

so that the difference of the variance between the two estimators becomes:

$$Var(\hat{d})-Var(\hat{d}_c)\approx\frac{1}{n\,o}\,[o-1]2\hat{S}^2\;-\frac{\rho}{n\,o}[o^2-1]2\hat{S}^2\;=$$

$$=\frac{2\hat{S}^2}{n\,o}(1-o)[\rho(1+o)-1]$$

Therefore:

1) if $\rho=\frac{1}{(1+o)}: Var(\hat{d})-Var(\hat{d}_c)=0$

   and the use of one estimator compared to the other one is indifferent

2) if $\rho>\frac{1}{(1+o)}: Var(\hat{d})-Var(\hat{d}_c)>0$
   and the estimator that use only the overlap between the two sample ($\hat{d}_c$) is better than the estimator $\hat{d}$

As we can see (Table 1.1), if the size of the overlap is considerable, then it is convenient to use the estimator $\hat{d}_c$, since a not too high value of the correlation is required.

Especially for economic surveys (e.g. the turnover survey), where the correlation between the observed variables over time is usually high, it is better to use the estimator based on the overlap ($\hat{d}_c$).

Table 1.1 When it is better to use the estimator $\widehat{d}_c$ for different overlapping rates

overlap	ρ
0,1	>0,91
0,2	>0,83
0,3	>0,77
0,4	>0,71
0,5	>0,67
0,6	>0,63
0,7	>0,59
0,8	>0,56
0,9	>0,53
1	>0,50

1.3 - The variance of change in dynamic non-stratified population

Laniel (1988) considered in his research, the case of the difference between levels in two consecutive occasion $d = Y_2 - Y_1$, removing the hypothesis of no birth-mortality in the population.

At the first occasion, Laniel considered a sample s_1 of size n_1 selected with srswor from the population U_1. Between the first and second occasion there are change in the population, due to the birth-mortality of the units. At the second occasion, Laniel identified the units in the population U_2 that have survived from the first occasion (U_c) and the new units (U_b) referring to births. Laniel also identified the units in the sample s_1 that survived at the second occasion. We will refer to them with s_{1c}, of random size n_{1c} (Figure 1.10).

The sample at the second occasion comes from two independent sampling performed on U_c and U_b respectively.

He distinguished the sampling from U_c in two cases, according to a modified version previously described in Tam.

1. In the sampling plan A, a part of the first sample, of size rn_{1c} (where $0 < r < 1$), is randomly selected from the sample s_{1c}, while the remaining part, of size $(1 - r)$ is selected according to srswor from $U_c - s_{1c}$ (see Figure 1.11).

2. In the sampling plan B, a part of the first sample, of size rn_{1c}, is randomly selected from the sample s_{1c}, while the remaining part, of size $(1 - r)n_{1c}$ is selected according to srswor from $U_c - rs_{1c}$ (see Figure 1.12).

Figure 1.10. Effect of birth-mortality in population on the sample

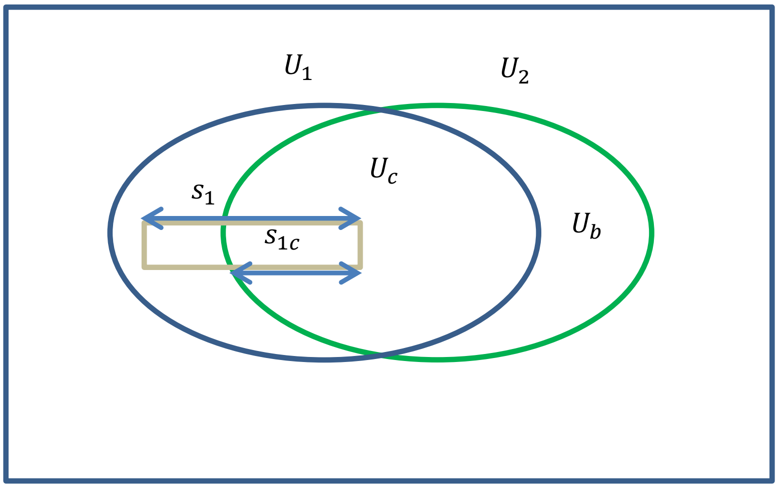


Figure 1.11. Sampling plan A

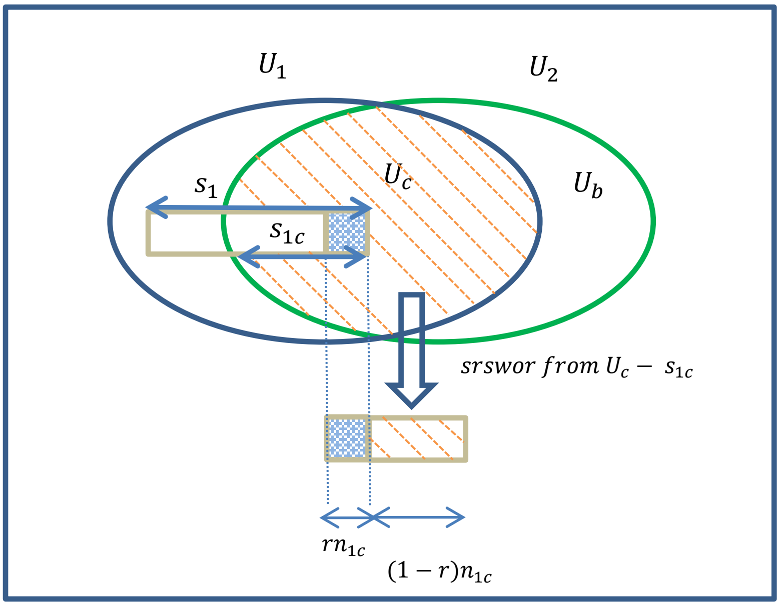
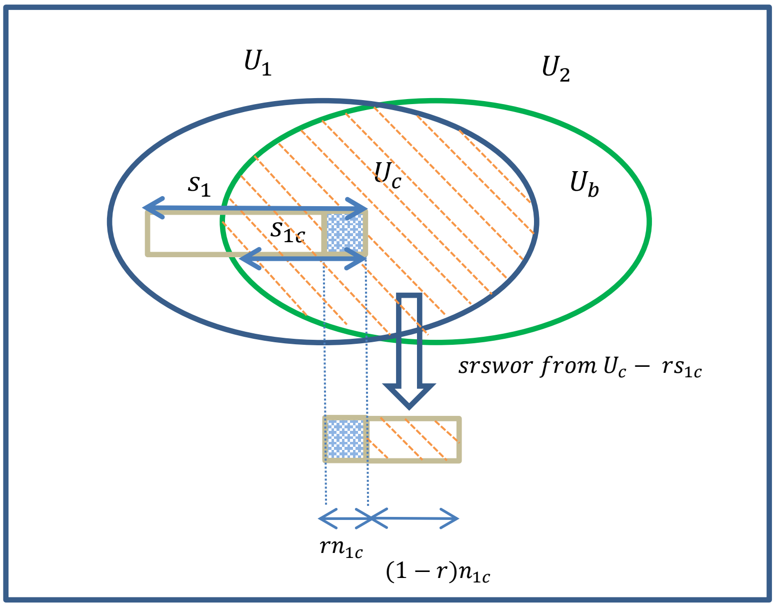


Figure 1.12. Sampling plan B



The estimates for the total level in the two different occasions can be estimated through the expansion estimator.

At the first occasion, the estimate for the total is:

$$\hat{Y}_1 = \frac{N_1}{n_1} \sum y_{1i} \,.$$

At the second occasion, the estimate of Y_2, is given by the sum of the total in U_c and the total in U_b:

$$\hat{Y}_2 = \hat{Y}_c + \hat{Y}_b$$

$$\hat{Y}_c = \frac{N_c}{n_{1c}} \sum y_{1i}$$

$$\hat{Y}_b = \frac{N_b}{n_b} \sum y_{1i}$$

where the expansion factor $\frac{N_c}{n_{1c}}$ is a random variable.

The formula of the variance of the difference between the two occasions is:

$$Var\left(\hat{Y}_2 - \hat{Y}_1\right) = Var\left(\hat{Y}_1\right) + Var\left(\hat{Y}_b\right) + Var\left(\hat{Y}_c\right) - 2Cov(\hat{Y}_1, \hat{Y}_c)$$

The variance for the total $\hat{Y}_1$ and $\hat{Y}_b$, are known to be (Cochran, 1977, p.23):

$$Var(\hat{Y}_1) = N_1^2(1-f_1)\frac{S_{y_1}^2}{n_1} = N_1^2\left(\frac{1}{n_1}-\frac{1}{N_1}\right)S_{y_1}^2,$$

$$Var(\hat{Y}_b) = N_b^2(1-f_b)\frac{S_{y_b}^2}{n_b} = N_b^2\left(\frac{1}{n_b}-\frac{1}{N_b}\right)S_{y_b}^2.$$

Using Lemma 1 of Tam (1984, p.288), Laniel found the formula of the variance for $\hat{Y}_c$

$$Var(\hat{Y}_c) = N_c^2\left(E\left[\frac{1}{n_{1c}}\right]-\frac{1}{N_c}\right)S_{y_c}^2\,,$$

where, following Sukhatme P. & Sukhatme B. (1970), and assuming n_1 sufficiently large,

$$E\left[\frac{1}{n_{1c}}\right] = \frac{N_1}{n_1 N_C}\left[1+\frac{(N_1-n_1)(N_1-N_c)}{(N_1-1)n_1N_c}\right].$$

Using Lemma 2 and Lemma 3 of Tam (1984, p.289), Laniel found the estimate for the covariance between $\hat{Y}_1, \hat{Y}_c$:

$$Cov(\hat{Y}_1,\hat{Y}_c) = \\ = \begin{cases} N_1N_c\left(\dfrac{r}{n_1}-\dfrac{E[n_{1c}]}{n_1N_c}\right)S_{y_1y_2}, & for\ Sampling\ Plan\ A \\ N_1N_c\left(\dfrac{r}{n_1}-\dfrac{E[n_{1c}]}{n_1N_c}+\dfrac{1-r}{n_1}E\left[\dfrac{n_{1c}}{N_c-rn_{1c}}\right]\right)S_{y_1y_2}, & for\ Sampling\ Plan\ B \end{cases}$$

where in the above expression, supposing $n_1$sufficiently large and using the second order Taylor's formula, we have

$$E\left[\frac{n_{1c}}{N_c-rn_{1c}}\right] = \frac{n_1}{N_1-rn_1}1+\frac{(N_1-n_1)(N_1-N_c)N_1rn_1}{(N_1-1)(N_1-rn_1)^2N_c}.$$

Laniel found the formula of the estimate of the variance of the change between two consecutive occasions in dynamic population, but he did not consider that in many repeated surveys, in particular business surveys (eg. the italian quarterly service turnover survey), a stratified simple random sample without replacement (stratified srswor) is actually used.

1.4 - The variance of change in dynamic stratified population

Tam (1984) and Qualité and Tillé (2008) do not provide an explicit form for stratification. However, under the assumption of a fixed population, fixed sample size and overlapping rate, and constant stratification over time, the result can be easily derived (Andersson, 2011):

$$\widehat{Cov}\,(\bar{y}_2, \bar{y}_1) = \sum_h \left(\frac{n_{h,c}}{n_{h,1}n_{h,2}} - \frac{1}{N_h}\right)\hat{S}_{h,\,y_1y_2} = \sum_h \left(\frac{o_{h,c}}{n_{h,2}} - \frac{1}{N_h}\right)\hat{S}_{h,y_1y_2},$$

where $\hat{\mathrm{S}}_{h,\mathrm{y}_1\mathrm{y}_2}$ is the covariance between y_1 and y_2 calculated on the common sample s_c within stratum h.

Knottnerus and Van Delden (2012) considered in their research the case of rotating panels and population with dynamic strata and change in the population over time. In this case, there are three aspects that must be taken into consideration.

1. As specified in Holt & Skinner (1989) and Kitagawa (1955) in the case of a stratified population, the net change in population means between two different occasions, $\Delta = \bar{Y}_2 - \bar{Y}_1$, can be decomposed as a sum of a component referred to the change in population mean, assuming no change in the stratum composition in the population between the two occasions (A), and the change in the stratum composition (due to different stratum classification of the same units in both occasion or births and deaths), assuming no change in the mean within stratum h (B):

$$\Delta = \bar{Y}_2 - \bar{Y}_1 = A + B$$
$$= \sum_{h=1}^{H} W_{1h}(\bar{y}_{2h} - \bar{y}_{1h}) + \sum_{h=1}^{H} \bar{Y}_{1h}(W_{2h} - W_{1h})$$

In this formulation, as we can see, in the first term the stratum composition is fixed at the first occasion and we only measure the change in the mean within stratum while in the second term the mean within stratum is fixed at the first occasion and we measure only the change in the stratum composition between the two occasions.

2. Between two occasions, some units in the population could change their stratification variables value. This is very common for business surveys, where

one of the stratification variables can be the number of employees, so a different number of employees between one occasion and another can lead some units to be classified into different stratum over time. Because of these stratum migration, some estimates referred to a stratum "h" on the first occasion, could be correlated with the estimates of the stratum "l" on the second occasion.

3. As the population is repeatedly sampled, sample overlap may occur between the two occasions, as already discussed in the previous paragraphs.

Considering these aspects, Knottnerus and Van Delden derives the formula of variance for the yearly relative grow rate of Ducht monthly turnover in Supermaket. The survey is based on a rotating sample stratified by economic activity and size. The sample is monthly updated to take into account of births and deaths in the population. The sample is updated also in January of every year: the 10% of the sample is replaced with other units and the units that remain in the sample are stratified according to their actual size.

Due to the migration of units from one stratum to another, strata are probably composed by units with different inclusion probabilities. To solve this problem the authors form substrata that take into account the reallocation of the units from the stratum h in December to the stratum l in January. They define the following quantities

- $U_{hl}^{dec,jan}$ = set of units in the population that have migrated from the stratum h in december to the stratum l in january, having size $N_{hl}^{dec,jam}$.
- $U_{0l}^{dec,jan}$ = set of births in the population is the stratum h, having size $N_{0l}^{dec,jan}$.

Let now $\hat{g}^{t,t-12}$ be the relative change of the monthly turnover with respect to the same month of the previoulsy year:

$$\hat{g}^{t,t-12} = \frac{\hat{Y}^{t}}{\hat{Y}^{t-12}} - 1\,,$$

and define further:

$$\hat{G}^{t,t-12} = \frac{\hat{Y}^{t}}{\hat{Y}^{t-12}} = 1 + \hat{g}^{t,t-12}$$

To estimate the variance of $\hat{g}^{t,t-12}$ the authors use the first order Taylor expansion of a ratio between two estimators, obtaining:

$$var(\hat{g}^{t,t-12}) = var\left(\hat{G}^{t,t-12}\right) =$$
$$= var\left(\frac{\hat{Y}^{t}}{\hat{Y}^{t-12}}\right) \approx \frac{var\left(\hat{Y}^{t} - G^{t,t-12}\,\hat{Y}^{t-12}\right)}{(Y^{t-12})^{2}} =$$
$$= \frac{var(\hat{Y}^{t}) + (G^{t,t-12})^{2} var(\hat{Y}^{t-12}) - 2G^{t,t-12} cov(\hat{Y}^{t-12}, \hat{Y}^{t})}{(Y^{t-12})^{2}}$$

Considering the sampling design, the covariance term can be written as

$$cov\left(\hat{Y}^{t-12}, \hat{Y}^{t}\right) = cov\left(\sum_{h=1}^{H} N_{h}^{t-12} \bar{y}_{h}{}^{t-12}, \sum_{l=1}^{H} N_{l}^{t} \bar{y}_{l}{}^{t}\right) =$$
$$\sum_{h=1}^{H}\sum_{i=1}^{H} N_{h}^{t-12} N_{l}^{t}\; cov(\bar{y}_{h}{}^{t-12}, \bar{y}_{l}{}^{t})$$

where $\bar{y}_{h}{}^{t}$ and $\bar{y}_{l}{}^{t-12}$ are the mean of the turnover in the sample in the stratum h in month t and in the stratum l and in the same month of the previously year, respectively.

To take into account the reallocation of the units within stratum, and proceed with the analysis, the authors define the variables:

- $N_{hl}^{t-12,t}$ = size of the units in the population that at the occasion t belong in the stratum h while in the same month of the previously year belonged in the stratum l ($U_{hl}^{t-12,t}$).
- n_{hl}^{t-m} = size of the units in the sample in the month $t-m$ ($m = 0 \leq m \leq 12$) within the substratum hl (s_{hl}^{t-m}): the size in s_{hl}^{t-m} can be different from that in s_{hl}^{t} because the units that belonged to s_{hl}^{t-m} and change stratum, can be not selected in t, thus no belong to s_{hl}^{t}
- Y_{hl}^{t-m} and $\bar{Y}_{hl}^{t-m}$ are the total and the mean turnover, respectvely, within population in month $t-m$ ($m = 0, 12$), of the units that have migrated from the stratum h to the stratum l in january ($U_{hl}^{t-12,t}$).
- y_{hl}^{t-m} and $\bar{y}_{hl}^{t-m}$ = total and the mean turnover, respectvely, within the sample in month $t-m$ ($m = 0, 12$), of the units that have migrated from the stratum h to the stratum l in the population, between t-12 and t.

- $n_{hl}^{t-12,t}$ = size of the overlapping units between the sample at the t and $t-12$ occasion within the substratum hl.
- $\bar{y}_{hl}^{t-m}$ = mean turnover in the month $t-m$ $(m=0,12)$ of the overlapping units, within the substratum hl $s_{hl}^{t-12,t}$

They define $h=(0,1,\dots.H)$ at the occasion t-12 and $l=(1,\dots.H,H+1)$ at the occasion t, to take into account the birth $(h=0)$ and the mortality $(l=H+1)$ respectively.

Then the $cov(y_h^{t-12}y_l^t)$ term can be formuled as

$$cov(\bar{y}_h^{t-12}\bar{y}_l^t)=cov\left(\sum_{g=1}^{H+1}\frac{n_{hg}^{t-12}}{n_h^{t-12}}\bar{y}_{hg}^{t-12},\sum_{k=0}^{H}\frac{n_{kl}^t}{n_l^t}\bar{y}_{kl}^t\right)$$

$$=\frac{1}{n_h^{t-12}n_l^t}cov\left(\sum_{g=1}^{H+1}n_{hg}^{t-12}\bar{y}_{hg}^{t-12},\sum_{k=0}^{H}n_{kl}^t\bar{y}_{kl}^t\right)$$

Since:

- $cov\left(n_{hg}^{t-12}\bar{y}_{hg}^{t-12},n_{kl}^t\bar{y}_{kl}^t\right)=0$ for $k\neq h$
- $cov\left(n_{hg}^{t-12}\bar{y}_{hg}^{t-12},n_{hl}^t\bar{y}_{hl}^t\right)=$

$$=E\,cov\left(n_{hg}^{t-12}\bar{y}_{hg}^{t-12},n_{hl}^t\bar{y}_{hl}^t|n_{hg}^{t-12},n_{hl}^t\right)$$
$$+\,cov\left\{E\left(n_{hg}^{t-12}\bar{y}_{hg}^{t-12}|n_{hg}^{t-12},n_{hl}^t\right),E\left(n_{hl}^t\bar{y}_{hl}^t|n_{hg}^{t-12},n_{hl}^t\right)\right\}=$$

$$=0+cov\left\{\,n_{hg}^{t-12}E\,\left(\bar{y}_{hg}^{t-12}|n_{hg}^{t-12},n_{hl}^t\right),\,n_{hg}^tE\,\left(\bar{y}_{hl}^t|n_{hg}^{t-12},n_{hl}^t\right)\right\}=$$

$$=cov\left(\,n_{hg}^{t-12}\bar{Y}_{hg}^{t-12},\,n_{hg}^t\bar{Y}_{hl}^t\right)=\bar{Y}_{hg}^{t-12}\bar{Y}_{hl}^t\,cov\left(\,n_{hg}^{t-12},\,n_{hg}^t\right)=0\ \ \textit{for}\ g\neq l\,,$$

we obtain:

$$cov(\bar{y}_h^{t-12}\bar{y}_l^t)=\frac{1}{n_h^{t-12}n_l^t}\,cov\,(n_{hl}^{t-12}\bar{y}_{hl}^{t-12},n_{hl}^t\bar{y}_{hl}^t).$$

Conditionally on $v_{hl}=\left(n_{hl}^{t-12},n_{hl}^{t-12,t},n_{hl}^t\right)$, the covariance term can be expressed in this way:

$$cov\,(n_{hl}^{t-12}\bar{y}_{hl}^{t-12},n_{hl}^t\bar{y}_{hl}^t)=$$
$$=E\,cov(n_{hl}^{t-12}\bar{y}_{hl}^{t-12},n_{hl}^t\bar{y}_{hl}^t|v_{hl})$$
$$+\,cov\,\{E(n_{hl}^{t-12}\bar{y}_{hl}^{t-12}|v_{hl}),E(n_{hl}^t\bar{y}_{hl}^t|v_{hl})\},$$

where the second term is equal to:

$$cov\,\{E(n_{hl}^{t-12}\bar{y}_{hl}^{t-12}|v_{hl}), E(n_{hl}^{t}\bar{y}_{hl}^{t}|v_{hl})\} = cov\{\,n_{hl}^{t-12}\;E(\bar{y}_{hl}^{t-12}), n_{hl}^{t}E(\bar{y}_{hl}^{t})\}$$
$$= \bar{Y}_{hl}^{t-12}\bar{Y}_{hl}^{t}\;cov(\,n_{hl}^{t-12},\;n_{hl}^{t}) = 0^{1}.$$

Developing the first term and postponing the various steps to the paper of Knottnerus and Van Delden, we obtain

$$cov(\bar{y}_h^{t-12}\bar{y}_l^{t}) = E\;\left\{\frac{n_{hl}^{t-12}n_{hl}^{t}}{n_h^{t-12}n_l^{t}}\left(\frac{n_{hl}^{t-12,t}}{n_{hl}^{t-12}n_{hl}^{t}} - \frac{1}{N_{hl}^{t-12,t}}\right)S_{hl}^{t-12,t}\right\}.$$

This expression can be estimated from the overlapping sample :

$$\widehat{cov}(\bar{y}_h^{t-12}\bar{y}_l^{t}) = \frac{n_{hl}^{t-12}n_{hl}^{t}}{n_h^{t-12}n_l^{t}}\left(\frac{n_{hl}^{t-12,t}}{n_{hl}^{t-12}n_{hl}^{t}} - \frac{1}{N_{hl}^{t-12,t}}\right)\hat{S}_{hl,OLP}^{t-12,t}\,,$$

where

$$\hat{S}_{hl,OLP}^{t-12,t} = \frac{1}{n_{hl}^{t-12,t}-1}\sum_{i=1}^{n_{hl}^{t-12,t}}\left(y_{hli}^{t-12} - \bar{y}_{hl,OLP}^{t-12}\right)\left(y_{hli}^{t} - \bar{y}_{hl,OLP}^{t}\right).$$

As $n_{hl}^{t-12,t}$ is sufficiently large, this is a reasonable estimate while for small value could lead to a negative estimates in the numerator of $var(\hat{g}^{t,t-12})$. For this reason, Knottnerus and Van Delden proposed an alternative estimator to $\hat{S}_{hl,OLP}^{t-12,t}$, namely

$$\hat{S}_{hl}^{t-12,t} = \hat{\rho}_{hl,Olp}^{t-12,t}\hat{S}_{hl}^{t-12}\hat{S}_{hl}^{t}\,,$$

where:

$$\hat{S}_{hl}^{t-m} = \sqrt{\frac{1}{n_{hl}^{t-m}-1}\sum_{i=1}^{n_{hl}^{t-m}}(y_{hli}^{t-m} - \bar{y}_{hl}^{t-m})^2} \qquad m = (0,12)$$

$\hat{\rho}_{hl,Olp}^{t-12,t}$is computed on the sample $s_{hl}^{t-12,t}$ and is the estimate of the correlation between y^{t} and y^{t-12} in $U_{hl}^{t-12,t}$.

With this estimator in the formula of the estimate of the covariance, they computed the confindence interval of the growth rates of the turnover for the ducht supermarket.

Summarizing the various steps, then we have:

[1] Like described by the authors, we will see that Nordberg (2000) derived a different result for this expression

$$\widehat{cov}\left(\hat{Y}^{t-12}, \hat{Y}^{t}\right) = \widehat{cov}\left(\sum_{h=1}^{H} N_h^{t-12} \bar{y}_h{}^{t-12}, \sum_{l=1}^{H} N_l^{t} \bar{y}_l{}^{t}\right) =$$

$$= \sum_{h=1}^{H} \sum_{i=1}^{H} N_h^{t-12} N_l^{t}\ \widehat{cov}(\bar{y}_h{}^{t-12}, \bar{y}_l{}^{t}) =$$

$$= \sum_{h=1}^{H} \sum_{i=1}^{H} N_h^{t-12} N_l^{t} \frac{1}{n_h^{t-12} n_l^{t}}\ \widehat{cov}\ (n_{hl}^{t-12} \bar{y}_{hl}^{t-12}, n_{hl}^{t} \bar{y}_{hl}^{t}) =$$

$$= \sum_{h=1}^{H} \sum_{i=1}^{H} N_h^{t-12} N_l^{t} \frac{n_{hl}^{t-12} n_{hl}^{t}}{n_h^{t-12} n_l^{t}} \left(\frac{n_{hl}^{t-12,t}}{n_{hl}^{t-12} n_{hl}^{t}} - \frac{1}{N_{hl}^{t-12,t}}\right) \hat{S}_{hl}^{t-12,t} =$$

$$= \sum_{h=1}^{H} \sum_{i=1}^{H} \frac{N_h^{t-12} N_l^{t}}{n_{hl}^{t-12} n_{hl}^{t}} n_{hl}^{t-12,t} \left(1 - \frac{n_{hl}^{t-12} n_{hl}^{t}}{n_h^{t-12} N_{hl}^{t-12,t}}\right) \hat{S}_{hl}^{t-12,t}$$

Nordberg (2000), using inclusion indicators, derived the formula for the variance of change over time, considering dynamic stratified populations, with units that can migrate between strata. He calculated this formula for the business survey in Statistics Sweden. To increase the precision of the estimates overt time, the sampling design in Statistics Sweden is constructed by using the Samu system[2]. This system is based on permanent random numbers associated with the units in the frame populations, and is used in particular to ensure a given overlapping between consecutive samples. A random number is associated with each units in the frame population at time 1 (frame1), then the frame is ordered by this number, and from a predeterminate starting point the first n_h units within the stratum h $(h = (0,1, \dots H))$ are selected. At time 2, there is a new updated frame population that take into account of birth-mortality on the enterprises (frame 2). Each unit in frame 2 that was also in the frame 1, maintains his permanent random number. At each new units a new random number is assigned, while units that were in frame 1 and no in the frame 2 are discarded. Then the units in the frame 2 are order on the basis of assigned random number, and from a starting point, the first n_l units within the stratum l $(l = (0,1, \dots L))$ are selected. To obtain the maximum overlap between the two samples, the starting point in frame 2 is the same used in frame 1. A random effect is due to the birth-mortality between the two populations over time.

[2] See "Samu: the system for co-ordination of frame populations and sample from the Business Register at Statistic Sweden", Background Facts on Economic Statistics (2003)

Nordberg considers the estimates of the totals at time 1 and a time 2, obtained with both the Horvitz-Thompson estimator (H-T) and the Generalised Regression estimator (GREG). Then the estimates for the variable at time 1 (Y_1) and the variable at time 2 (Y_2) are functions of these totals. For istance, using the Horvitz Thompson estimator we obtain:

$$\hat{t}_{1,j} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in U_1} y_{1,j,k} \delta_{1,k}$$

$$\hat{t}_{2,j} = \sum_{l=1}^{L} \frac{N_l}{n_l} \sum_{r \in U_1} y_{2,j,r} \delta_{2,r}$$

where:

$$\delta_{1,k} = \begin{cases} 1, & k \in s_1 \\ 0, & otherwise \end{cases}$$

$$\delta_{2,l} = \begin{cases} 1, & l \in s_2 \\ 0, & otherwise \end{cases}$$

The estimates of $Y_1$and Y_2 are then obtained as functions of totals $\hat{t}_{1,j}$ and $\hat{t}_{2,j}$ respectively:

$$\hat{Y}_1 = f(\hat{t}_{1,1}, \hat{t}_{1,2}, \dots, \hat{t}_{1,J})$$

$$\hat{Y}_2 = f(\hat{t}_{2,1}, \hat{t}_{2,2}, \dots, \hat{t}_{2,J})$$

The formula of the covariance in the expression of the variance of change $\hat{d}$, can be written using a Taylor linearisation:

$$C(\hat{Y}_1, \hat{Y}_2) \approx \sum_i \sum_j f_i'(\hat{t}_{1,1}, \hat{t}_{1,2}, \dots, \hat{t}_{1,J}) f_j'(\hat{t}_{2,1}, \hat{t}_{2,2}, \dots, \hat{t}_{2,J}) C(\hat{t}_{1i}, \hat{t}_{2j}),$$

where $f_i' = \partial f / \partial t_{1i}$, $f_j' = \partial f / \partial t_{2i}$ and $C(\hat{t}_i, \hat{t}_j)$ is the covariance between $\hat{t}_{1i}, \hat{t}_{2j}$.

The units in the two populations can be split into death (D), overlapping (O) and born units (B).

Since the population is stratified, we can split the three groups D, O and B in D_h, O_{hl}, B_l, where:

- D_h is the subset of units belonging only to the frame 1, within stratum h. Its size is G_{h+}.

- O_{hl} is the subset of the overlapping units between the two frame, within stratum h in frame 1, and in stratum l in frame 2. Its size is G_{hl}.
- B_l is the subset of units belonging only to the frame 2, within stratum l. Its size is G_{+l}.

n_h and n_l are the size of the sample within stratum h in the frame 1 and of the sample within stratum l in the frame 2, respectively. They can be calculated as:

$$n_h = a_{h+} + \sum_{l=1}^{L} a_{hl} \, ,$$

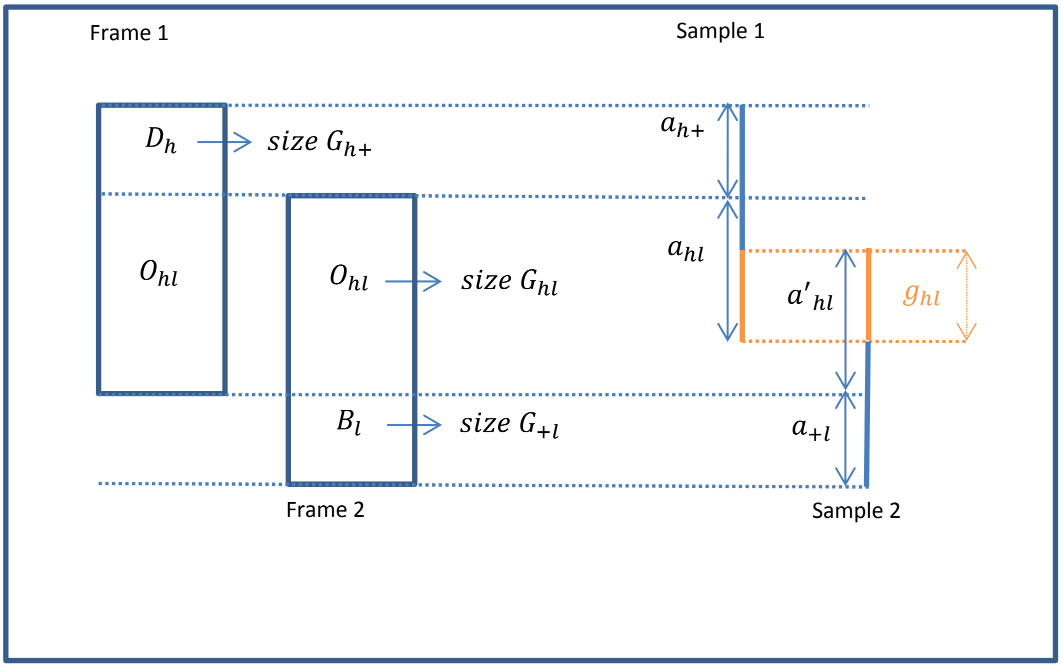
$$n_l = a_{+l} + \sum_{h=1}^{H} a'_{hl} \, ,$$

where a_{h+} and a_{+l} are the size of the sample units belonging to the group D_h and B_l respectively, while a_{hl} and a'_{hl} are the size of the sample units at time 1 and at time 2 respectively, belonging to the group O_{hl}. Furthermore, let g_{hl} be the size of the sample overlapping units, belonging to the group O_{hl}. The rappresentation just described is showed in Figure 1.13.

Nordberg use the random quantity $\Omega = \{a_{hl}, a'_{hl}, g_{hl}, a_{h+}, a_{+l}\} \, \forall \, h = 1, \dots, H; l = 1, \dots., L$, to split $C(\hat{Y}_1, \hat{Y}_2)$ in:

$$C(\hat{Y}_1, \hat{Y}_2) = E_\Omega\left(C(\hat{Y}_1, \hat{Y}_2|\Omega)\right) + C_\Omega\left(E(\hat{Y}_1|\Omega), E(\hat{Y}_2|\Omega)\right)$$

Figure 1.13. Frame population and sample units between the occasion 1 and 2



Let's see now, how Nordberg calculates the two terms:

1) The covariance in the first term $E_\Omega\left(C(\hat{Y}_1, \hat{Y}_2|\Omega)\right)$, can be calculated as:

$$C(\hat{Y}_1, \hat{Y}_2|\Omega) \approx \sum_i \sum_j f_i'(\hat{t}_{1,1}, \hat{t}_{1,2}, \dots, \hat{t}_{1,J}) f_j'(\hat{t}_{2,1}, \hat{t}_{2,2}, \dots, \hat{t}_{2,J}) C(\hat{t}_{1i}, \hat{t}_{2j}|\Omega)\,.$$

Using the expression of the H-T estimators, we obtain:

$$C(\hat{t}_{1i}, \hat{t}_{2j}|\Omega) = \sum_{h=1}^{H} \sum_{l=1}^{L} \sum_{k \epsilon U_h} \sum_{r \epsilon U_l'} \frac{N_h N_l'}{n_h n_l'} y_{1,i,k} y_{2,j,r} C(\delta_{1,k}, \delta_{2,r}|\Omega) =$$

$$= \sum_{h=1}^{H} \sum_{l=1}^{L} \sum_{k \epsilon U_h} \sum_{r \epsilon U_l'} \frac{N_h N_l'}{n_h n_l'} y_{1,i,k} y_{2,j,r} (E(\delta_{1,k}, \delta_{2,r}|\Omega) - E(\delta_{1,k}|\Omega) E(\delta_{2,r}|\Omega))\,.$$

The unbiased estimator for $C(\hat{t}_{1i}, \hat{t}_{2j}|\Omega)$ is:

$$\hat{C}(\hat{t}_{1i}, \hat{t}_{2j}|\Omega) = \sum_{h=1}^{H} \sum_{l=1}^{L} \sum_{k \epsilon U_h} \sum_{r \epsilon U_l'} \frac{N_h N_l'}{n_h n_l'} y_{1,i,k} y_{2,j,r} \left(1 - \frac{E(\delta_{1,k}|\Omega) \mathrm{E}(\delta_{2,r}|\Omega)}{E(\delta_{1,k}, \delta_{2,r}|\Omega)}\right) \delta_{1,k} \delta_{2,r}\,.$$

Nordberg also computed the first and second order inclusion probabilietis, and obtained:

$$\hat{C}(\hat{t}_{1i}, \hat{t}_{2j}|\Omega) = \sum_{h=1}^{H} \sum_{l=1}^{L} \frac{N_h N_l' \tilde{a}_{hl}}{G_{hl}^2 n_h n_l'} \frac{G_{hl}(G_{hl} - \tilde{a}_{hl})}{(\tilde{a}_{hl} - 1)} \left\{ \sum_{k \in P_{hl}} y_{1,i,k} y_{2,j,r}\, \delta_{1,k} \delta_{2,k} - \frac{1}{\tilde{a}_{hl}} \left(\sum_{k \in P_{hl}} y_{1,i,k}\, \delta_{1,k} \right) \left(\sum_{r \in P_{hl}} y_{2,j,r} \delta_{2,r} \right) \right\},$$

where $\tilde{a}_{hl} = \frac{a_{hl} a_{hl}'}{g_h}$.

Hence:

$$\hat{C}(\hat{Y}_1, \hat{Y}_2|\Omega) \approx \sum_i \sum_j f_i'(\hat{t}_{1,1}, \hat{t}_{1,2}, \dots, \hat{t}_{1,J}) f_j'(\hat{t}_{2,1}, \hat{t}_{2,2}, \dots, \hat{t}_{2,J}) \hat{C}(\hat{t}_{1i}, \hat{t}_{2j}|\Omega)$$

2) Nordberg considers the second term as a remainder term. He proposed a method to calculate it for the Swedish sampling disegn, trought a computer intensive

procedure with the Samu system. This term in Knottnerus and Val Deldend is instead estimated to be 0 (see again above). Nordberg estimates this term in this way:

$$C_{\Omega}\left(E\left(\hat{Y}_1|\Omega\right),E\left(\hat{Y}_2|\Omega\right)\right)=$$
$$=\sum_i\sum_j f_i'\left(\hat{t}_{1,1},\hat{t}_{1,2},\dots,\hat{t}_{1,J}\right)f_j'\left(\hat{t}_{2,1},\hat{t}_{2,2},\dots,\hat{t}_{2,J}\right)C_{\Omega}\left(E(\hat{t}_{1i}|\Omega),E(\hat{t}_{2j}|\Omega)\right)$$

where:

$$\begin{aligned}E(\hat{t}_{1i}|\Omega)&=\left(\sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in U_1}y_{1,i,k}\;E\left(\delta_{1,k}|a_{hl},a'_{hl},g_{hl},a_{h+},a_{+l}\right)\right)\\&=\sum_{h=1}^{H}\frac{N_h}{n_h}\left\{\left(\frac{a_{h+}}{G_{h+}}\sum_{k\in D_h}y_{1,i,k}\right)+\left(\sum_{l=1}^{L}\frac{a_{hl}}{G_{hl}}\sum_{k\in O_{hl}}y_{1,i,k}\right)\right\}\\&=\sum_{h=1}^{H}\frac{N_h}{n_h}\left\{\left(a_{h+}\sum_{k\in D_h}\frac{Y_{1,i,k}}{G_{h+}}\right)+\left(\sum_{l=1}^{L}a_{hl}\sum_{k\in O_{hl}}\frac{Y_{1,i,k}}{G_{hl}}\right)\right\}\\&=\sum_{h=1}^{H}\frac{N_h}{n_h}\left\{\left(a_{h+}\bar{Y}_{1,i,h+}\right)+\left(\sum_{l=1}^{L}a_{hl}\bar{Y}_{1,i,hl}\right)\right\},\end{aligned}$$

and similarly

$$\begin{aligned}E(\hat{t}_{2j}|\Omega)&=\left(\sum_{h=1}^{H}\frac{N_l}{n_l}\sum_{k\in U_2}y_{2,j,k}\;E\left(\delta_{2,k}|a_{hl},a'_{hl},g_{hl},a_{h+},a_{+l}\right)\right)\\&=\sum_{l=1}^{L}\frac{N_l}{n_l}\left\{\left(\frac{a_{+l}}{G_{+l}}\sum_{k\in B_l}y_{2,j,k}\right)+\left(\sum_{h=1}^{H}\frac{a'_{hl}}{G_{hl}}\sum_{k\in O_{hl}}y_{2,j,k}\right)\right\}=\\&=\sum_{l=1}^{L}\frac{N_l}{n_l}\left\{\left(a_{+l}\bar{Y}_{2,j,+l}\right)+\left(\sum_{h=1}^{H}a'_{hl}\bar{Y}_{2,j,hl}\right)\right\}.\end{aligned}$$

Then, to estimate $C_{\Omega}\left(E\left(\hat{Y}_1|\Omega\right),E\left(\hat{Y}_2|\Omega\right)\right)$ he applies the following procedure. He assigns a random number to each unit in the union of the two populations. Then such units are ordered by their random numbers, and the value $v_h=1$ is assigned to the firsts n_h units within stratum h of the first population, and the value $v_h'=1$ is assigned to the firsts n_l ordered units within stratum l of the second population.

To the other units the valued $v'_h = 0$ is assigned. Then, the following quantities are computed:

$$a_{h+}(m) = \sum_{k\in D_h} v_k(m), \qquad a_{hl}(m) = \sum_{k\in P_{hl}} v_k(m),$$

$$a'_{hl}(m) = \sum_{k\in P_{hl}} v'_k(m), \qquad a_{hl}(m) = \sum_{k\in P_{hl}} v'_k(m),$$

$$\hat{u}_{1,i}(m) = \sum_{h=1}^{H} \frac{N_h}{n_h}\left\{a_{h+}(m)\bar{y}_{1,i,h+} + \left(\sum_{l=1}^{L} a_{hl}(m)\,\bar{y}_{1,i,hl}\right)\right\}$$

$$\hat{u}_{2,j}(m) = \sum_{l=1}^{L}\frac{N_l}{n_l}\left\{\left(a_{+l}(m)\bar{y}_{2,j,+l}\right) + \left(\sum_{h=1}^{H} a'_{hl}(m)\bar{y}_{2,j,hl}\right)\right\}$$

These values are calculated for $m = 1, \ldots.,1000$. $\bar{y}_{1,i,h+}, \bar{y}_{1,i,hl}, \bar{y}_{2,j,+l}, \bar{y}_{2,j,hl}$ are the sample means associated to $\bar{Y}_{1,i,h+}, \bar{Y}_{1,i,hl}, \bar{Y}_{2,j,+l}, \bar{Y}_{2,j,hl}$. We can now calculate the estimates for the covariance $C_\Omega\left(E(\hat{Y}_1|\Omega), E(\hat{Y}_2|\Omega)\right)$, by:

$$\tilde{C}_{ij} = \frac{1}{1000}\sum_{m=1}^{1000}\left(\hat{u}_{1,i}(m)\hat{u}_{2,j}(m)\right) - \frac{\left(\sum_{m=1}^{1000}\hat{u}_{1,i}(m)\right)\left(\sum_{m=1}^{1000}\hat{u}_{2,j}(m)\right)}{1000}$$

and then:

$$\hat{C}_\Omega\left(E(\hat{Y}_1|\Omega), E(\hat{Y}_2|\Omega)\right) = \sum_i\sum_j f_i'(\hat{t}_{1,1},\hat{t}_{1,2},\ldots,\hat{t}_{1,J})f_j'(\hat{t}_{2,1},\hat{t}_{2,2},\ldots,\hat{t}_{2,J})\tilde{C}_{ij}.$$

1.5 - Other approaches

Berger (2004) also proposes a design-based estimator for covariance matrix that is adapted to overlapping samples between one wave and the next one, and he generalizes his results for stratified sample. He shows that his approach "yields non-negative definite estimates for covariance matrices and therefore positive variance estimates for a large class of measures of change".

Berger, based his results on the aggregation of conditional covariances, using a Poisson sampling approximation of the actual sampling scheme. While Hajeck (1964) developed his approach for a single sample, Berger extends this approach to overlapping samples. His assumptions are "a fixed number of units rotating in and out as well as a fixed number of units in the matched sample". These assumptions "hold with most rotating sampling scheme". However, as mentioned by Wood (2008) this method "involved a variety of matrix operations and no explicit covariance formula were presented".

Osier & Raymond (2017) describe possible approach to estimates the variance for annual changes in the European Union Labour Force Survey (EU-LFS) based indicators. Almost all the countries of the European Union use a 2-(2)-2 rotating design: the units in the sample are interviewed for two consecutive quarters, then leave for two quarters and return in the sample for two more quarters of the following year. Therefore, they have to take into account the overlap between quarterly and annual data. They suggest to adopt an estimator proposed by Berger & Priam (2013) and Berger & Oguz Alper (2015). This estimator can be used with several EU-LFS sampling designs, and is easy to implement because it does not require the calculation of the joint inclusion probability, that can be unknown with rotating designs. It can be implemented by standard statistical software as R, SAS, SPSS, Stata, and requires minimal computing power. The idea is to estimate the design covariance matrix of $\hat{y}_2$ and $\hat{y}_1$ ($\Sigma_{\hat{y}}$) in:

$$\widehat{Var}(\hat{d}) = \widehat{Var}(\hat{y}_2) + \widehat{Var}(\hat{y}_1) - 2\rho\left\{\widehat{Var}(\hat{y}_2)\widehat{Var}(\hat{y}_1)\right\}^{1/2} = \nabla^T \hat{\Sigma}_{\hat{y}} \nabla ,$$

from the covariance of the residuals ϵ_{1i} and ϵ_{2i} of the following multivariate linear regression model:

$$\begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} = \sum_{h=1}^{H} \begin{pmatrix} \beta_h^{(1)} z_{1h,i} + \beta_h^{(2)} z_{2h,i} + \beta_h^{(12)} z_{1h,i} z_{2h,i} \\ \gamma_h^{(1)} z_{1h,i} + \gamma_h^{(2)} z_{2h,i} + \gamma_h^{(12)} z_{1h,i} z_{2h,i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix},$$

where $\nabla = (-1,1)^T$, $i \epsilon s = s_1 \cup s_2$ and the residuals $(\epsilon_{1i}, \epsilon_{2i})$ have a bivariate distribution with null mean and unknown variance-covariance matrix. The covariate $z_{1h,i}$ and $z_{2h,i}$ are dummy design variables defined by:

$$z_{1h,i} = \begin{cases} 1, & if\ i \epsilon s_{1h} \\ 0, & otherwise \end{cases} \qquad z_{2h,i} = \begin{cases} 1, & if\ i \epsilon s_{2h} \\ 0, & otherwise \end{cases}$$

The $z_{1h,i}z_{2h,i}$ term, represents the interaction in the regression and take the rotation of the design into account. The $\beta_h^{(1)}, \beta_h^{(2)}, \beta_h^{(12)}, \gamma_h^{(1)}, \gamma_h^{(2)}, \gamma_h^{(12)}$ terms are the regression parameters of the model.

The model relies on the assumption that the sampling fractions are negligible, that is common thing for social survey like LFS. When we have large sampling fractions, which are common for example in business surveys, this approach is not suitable. Moreover, the estimator of the covariance matrix is unbiased only in the case of a large entropy.

One of the advantages of this approach is that the covariance matrix $\Sigma_{\hat{y}}$ is estimating using a single model, also if we have many stratum and totals. Moreover, in the case of the complex measures of the change ($\hat{d} = \hat{\theta}_2 - \hat{\theta}_1$ or $\hat{d} = \frac{\hat{\theta}_2}{\hat{\theta}_1}$, where $\hat{\theta}_2$ and $\hat{\theta}_1$ are smooth functions of estimators of totals $\hat{y}_2$ and $\hat{y}_1$), using Taylor linearization we have that:

$$\widehat{Var}(\hat{d}) = \nabla(\hat{y})^T \hat{\Sigma}_{\hat{y}}^{(A)} \nabla(\hat{y}),$$

where ∇(y) is the gradient of f(y) and the same estimated variance-covariance matrix $\Sigma_{\hat{y}}$ can be used for several measures of change (any function of the same totals). Therefore, the user, known the covariance matrix, have to define only the gradient. The estimate is possible without knowing the design and auxiliary variables because only the covariance matrix is necessary.

CHAPTER 2

Evaluation of the variance for the growth rate estimators used in the Istat service turnover survey

2.1 - Description of the survey and sampling design

The quarterly service turnover survey measures the quarterly percentage change recorded in sales at current prices by enterprises belonging to the domain of services (sections G, H, I, J, M, N of the Nace Rev. 2 classification), except for retail sales (G47). The indices are aggregated according to the Laspeyres formula, using a fixed weight structure that reflects the sectorial distribution of services turnover in the base year (figure 2.1). The quarterly service turnover index is obtained by aggregating all estimation domains.

The indicators produced up to March 2012 (G452, G46, H50, H51, H53, J) represented 60,1% of the total service turnover[3]. Istat's strategic aim for the period 2010-2013 has been to complete the set of indices for the services sector as required by European Regulation (Regulation No 1158/05 of the European Parliament and of the Council, annex D). For the quarterly turnover indices, this implied the creation of new surveys to increase the coverage of the indices already produced for other economic activities.

The planning and launch of the new surveys allowed in March 2012 the dissemination of the indices for the sectors G45-G452, H49, H52, I55, I56, reaching 84,9% of the total service turnover and the completion of the indices for the G, H and I sections. Moreover, in 2013 the launch of new surveys related to M and N sections allowed to complete the total set of indicators required (for more details see Bacchini et al. 2015).

[3] according the turnover weights structure of 2010

The turnover data are collected by a sample survey of about 17.000 enterprises. For the sectors where the market dynamics are determined by a small number of large companies (H50, H51, H53, J61, N78 domains of the Nace Rev. 2 classification), cut-off unit selection scheme have been adopted. In this case, the sample includes the biggest companies up to cover a sufficiently high share of the total turnover of the sector[4] (usually over 80%).

Table 2.1- The weights structure in 2015 for the quarterly turnover indicators of services

Nace Rev. 2	**Economic Activities**	**Weights 2015**
G45-G452	Wholesale & retail trade of motor vehicles and wholesale & retail trade and repair of motorcycles	8.792
G452	Maintenance and repair of motor vehicles	1.168
G46	Wholesale trade, except of motor vehicles and motorcycles	46.292
H49	Land transport and transport via pipelines	5.735
H50	Water transport	1.049
H51	Air transport	0.911
H52	Warehousing and support activities for transportation	4.752
H53	postal and courier activities	0.509
I 55	Accomodation	1.977
I 56	Food and beverage service activities	4.704
J	Information and comunication	9.237
M69	Legal and accounting activities	2.853
M70.2	management consultancy activities	1.240
M71	Architectural and engineering anctivities; technical testing and analysis	2.097
M73	Advertising and market research	1.112
M74	Other professional, scientific and technical activities	1.304
N78	Employment activities	0.771
N79	Travel agency, tour operator and other reservation service and related activities	0.992
N80	Security and investigation activities	0.325
N81.2	Cleaning activities	1.150
N82	Office administrative, office support and other business support activities	3.030
Total		**100.000**

[4] More information are available on the methodological note (istat.it)

However, for most sectors a stratified simple random sampling without replacement (stratified srswor) is used. The stratification variables are the economic activity and the size of the enterprise. Businesses above a given size threshold (usually 100 employees) are included in self-representative strata. For some sectors, a specific size threshold (usually of at least 2 employees) is applied in the sampling selection of companies.

Every year, the sample size is computed by means of the Bethel algorithms implemented in Mauss-R (see Barcaroli et al. 2010). This allows to minimize the sample size, given the maximum expected sampling errors on target estimates for each type of domain[5]. Estimation domains are the sub-populations at which level you want to compute the estimates of the parameters of interest. The precision required for the estimates, indicates the degree of reliability that the estimates have to guarantee. It is expressed in terms of the coefficient of variation (ratio between the standard error of the estimate and the estimate itself), to be specified for each parameter and each type of domain. The planned coefficient of variation for each estimation domain is fixed at 3%. The estimation domains are usually the 2 or 3 digits of the Nace Rev. 2 classification.

The auxiliary variables necessary for the allocation are stratification variables, that are essential to define strata and study domains, and the variables correlated with the variables of interest, useful for the study of their variability. The auxiliary information for the planning of the design is contained in the Istat Statistic Register of Active Firms (ASIA). The Register consists of economic units that run an activity in industrial and commercial sectors, as well as services to businesses and families sector. It provides identifying information (name and address) and business specific information (economic activity, employee number, activity start and end date, annual turnover) of these units[6]. The Register is annually updated through a process of integration of information from both administrative sources and statistical sources. Its regular maintenance guarantees the update of the complex of active economic units over time, ensuring an official data source, harmonized at European level, on the structure of the population of enterprises and on its demographic characteristics. The Register (also used for the calculation of the national accounts estimates) has a central role in the field of economic statistics: it identifies the reference population for the sampling plans and for the carryover to the universe of the main surveys on companies conducted by Istat.

[5] See the “User and methodological manual” about Mauss-R

[6] Istat.it -Schede standard di qualità -Archivio Statistico delle Imprese Attive

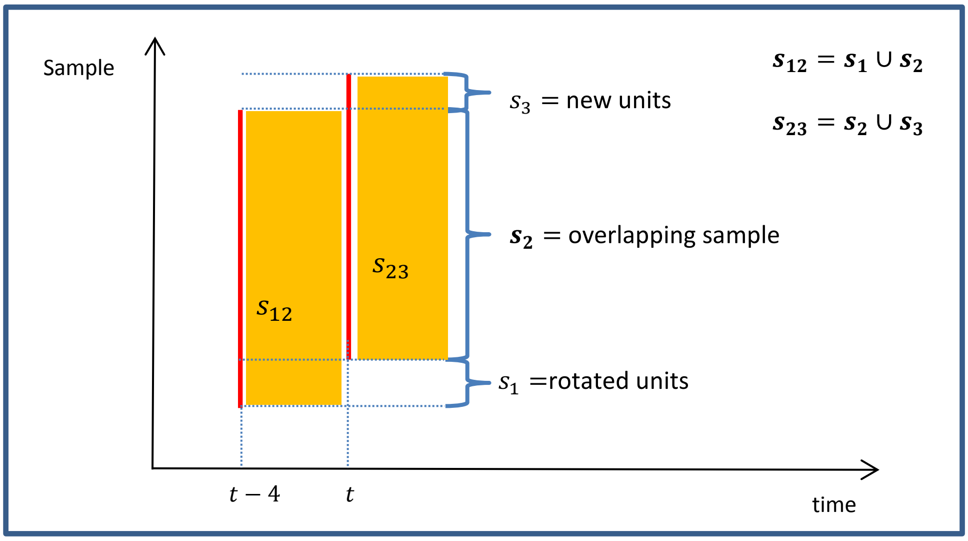
The latest available ASIA contains a delay of two years. This means that for the year 2019, the latest Asia contains information updated to 2017. For this reason, within the service turnover survey an integration with sample data is used. In particular, the annual turnover sample values obtained from the quarterly observations and the number of employees replace the values contained in Asia. Due to the high correlation with the quarterly turnover, the variable used to study of the variability of the variables of interest is the annual turnover.

The sample is updated to account for both a re-stratification of the units and a sample replacement of approximately 15%. The units in the sample are re-stratified according to their actual size and economic activity from Asia. Dead companies are discarded from the sample, together with the companies that have been in the sample for several years. New companies are randomly selected from the last Asia available excluding the units already in the sample (plan A of Tam), until the theoretical size provided by the Mauss-R software is reached within each stratum. In this way between two consecutive years we have two overlapping samples.

The situation just described is represented in Figure 2.1. As we can see, between one quarter (*t*) and the same quarters of the previous year (*t*-4) we have two different partially overlapping samples, s_{12} and s_{23}. The overlapping sample is represented by s_2.

The overlapping units (s_2) in the samples s_{12} and s_{23} could belong to different stratum, because the stratification variable values can change across the two consecutive years.

Figure 2.1. Overlapping sample between two consecutive years

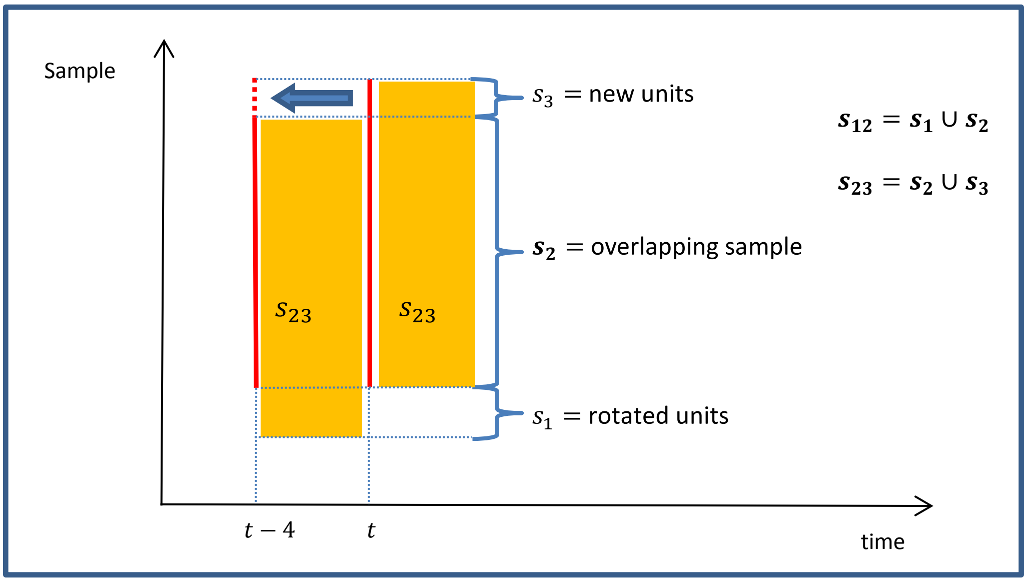


New companies entering in the sample (s_3) are required to indicate the turnover data for both the current year (*t*) and the previous year (*t-4*). In this way, it is possible to have turnover data for both estimation quarters, even if the firm was not in the sample s_{12} at the occasion *t-4*.

The estimates of the change between the occasion *t* and the occasion *t-4* are both computed on the sample s_{23}. It means that all observations are stratified in the same way over the two estimation quarters, according to the latest information available on the stratification variables. The rotated units are not included in the estimates, neither in the quarter *t* nor in the quarter *t-4*.

The situation is shown in Figure 2.2. The dashed red line indicates data referred to the *t-4* occasion that have been collected from the new enterprises entered in the sample at the occasion *t* (s_3).

Figure 2.2. Use of the new sample s_{23} for both quarters of estimation



2.2 - The methodology used for the growth rate estimation

As we have seen in the previous paragraph, the aim of the quarterly service turnover survey is the estimation of the percentage change of the turnover between the occasion *t* and *t-4*:

$$g = \left(\frac{Y_t}{Y_{t-4}} - 1\right)100 = (G-1)100$$

Let r_1 be the set of the respondent enterprises only at the occasion *t-4*, r_2 the set of respondent enterprises on both occasions *t-4* and *t*, r_3 the set of respondent enterprises only at the occasion *t*. Then we define $r_{12} = r_1 \cup r_2$ and $r_{23} = r_2 \cup r_3$.

The completion of the indicators for all the service sectors represented an opportunity to review the estimation procedure for G. For the new sectors, it is used a new estimation method that is different from the one used for the sectors already disseminated. In this section we analyze the two different methodologies.

2.2.1 - The estimator used for the sectors already disseminated

The estimation procedure for the sectors already disseminated before completing the indicator for all the service sectors is based on the variation computed on the overlapping sample units (olp) in both quarters. This means that only units in the sample s_{23} that respond in both quarters are directly involved in the estimate. The calculation of the change (G) is carried out at the stratum level. In formula, we can write:

$$\hat{G}_{h,olp} = \frac{\hat{\bar{Y}}^{t}}{\hat{\bar{Y}}^{t-4}} = \frac{\frac{1}{n_{r_2}} \sum_{k \in r_2} y_{h,k}^{t}}{\frac{1}{n_{r_2}} \sum_{k \in r_2} y_{h,k}^{t-4}} = \frac{\hat{\bar{y}}_{h,r2}^{t}}{\hat{\bar{y}}_{h,r2}^{t-4}}$$

By applying $\hat{G}_{olp}$ to the index number of the same stratum of the previous year, the stratum index for the current quarter is obtained:

$$\hat{I}_h^t = \hat{G}_{h,olp}\, I_h^{t-4}$$

The elementary stratum index consists of two parts: the first one is the ratio between the two sampling averages of turnover at the current occasion and at the occasion *t-4*, calculated on the set of common respondents r_2 within the stratum h ($\hat{G}_{h,olp}$). The second one is the published final stratum index for the same quarter of the previous year. The second part takes into account the change in the average level of the turnover for the quarter *t-4* compared to the same quarter of the base year. The index numbers are built in such a way that the average is equal to 100 in the base year.

The indices at the domain level are obtained by aggregating the stratum indices with an annual fixed weights system calculated via Asia, which takes into account the weights of the strata within the estimation domain, in terms of turnover:

$$\hat{I}_d = \sum_{h=1}^{H} \hat{I}_h w_h$$

Finally, the turnover change at the domain level can be calculated as follows:

$$\hat{G}_{d,olp} = \frac{\hat{I}_d^t}{I_d^{t-4}} = \frac{1}{I_d^{t-4}} \sum_{h=1}^{H} \hat{I}_h^t w_h =$$

$$\frac{1}{I_d^{t-4}} \sum_{h=1}^{H} \hat{G}_{h,olp}\ I_h^{t-4} w_h =$$

$$\frac{1}{I_d^{t-4}} \sum_{h=1}^{H} \frac{\hat{\bar{y}}_{h,r2}^{t}}{\hat{\bar{y}}_{h,r2}^{t-4}}\ I_h^{t-4} w_h$$

2.2.2 - The estimator used for the new sectors

For the new sectors, instead, a methodology for the estimation of the totals in the population has been adopted, which is based on all respondent enterprises in the two occasions (*all*). At the beginning the Horvitz-Thompson estimator has been computed and then a calibration estimator. Therefore, the initial sample weights are corrected using an auxiliary variable to account for non-response (Bacchini et al. 2014).

The change estimation at the stratum level through calibration is obtained by the ratio of the totals calculated within the stratum h:

$$\hat{G}_{h,all.cal} = \frac{\hat{Y}_{h,r_{23}}^{t}}{\hat{Y}_{h,r_{12}}^{t-4}} = \frac{\sum_{j \in r_{23}} y_{h,j}^{t} d_j}{\sum_{i \in r_{12}} y_{h,i}^{t-4}\ d_i}$$

where d_j and d_i are the calibration weights associated with the j-th unit and i-th unit respectively. The calibrated weights (d_j and d_i) associated with the same unit on the two survey occasions of investigation (*t* and *t-4*) can be different due to the different non-response on the two occasions (the sets of respondent enterprises r_{12} and r_{23} usually are not the same).

By summing up the totals of strata at the two occasions *t* and *t-4,* is possible to obtain the change estimation at the domain level:

$$\hat{G}_{d,all.cal} = \frac{\sum_{h=1}^{H} \hat{Y}_{h,r_{23}}^{t}}{\sum_{h=1}^{H} \hat{Y}_{h,r_{12}}^{t-4}} = \frac{\sum_{h=1}^{H} \sum_{j \in r_{23}} y_{h,j}^{t} d_j}{\sum_{h=1}^{H} \sum_{i \in r_{12}} y_{h,i}^{t-4} d_i}$$

The calibration variable used is the annual turnover, due to its high correlation with the variable of interest. The values of the calibration variable and the known totals are the same in both the numerator and the denominator, and derive from the latest available Asia together with integration on sample data. Calibration is performed at single stratum level, i.e. the known totals are calculated for each stratum. The estimated totals for each stratum are aggregated within the estimation domains to allow the calculation of $\hat{G}_{d,all.cal}$. By applying $\hat{G}_{d,all.cal}$ to the index number of the same estimation domain for the previous year, the domain index for the current quarter is obtained as follows:

$$\hat{I}_d^t = \hat{G}_{d,all.cal} * I_d^{t-4}$$

The estimation methodology adopted for the new sectors has some advantages with respect to the one used for the sector already disseminated. It includes in the calculation of the index all respondent companies, and not only the overlapping observations, like in the estimator $\hat{G}_{olp}$. In addition, the calibration can be implemented using the software ReGenesees (R Evolved Generalized Software for Sampling Estimates and Errors in Surveys)[7]. This software is a full-fledged R software for design-based and model-assisted analysis of complex sample surveys. This system is the outcome of a long-term research and development project aimed at defining a new Istat standard for calibration, estimation and sampling error assessment in large-scale sample surveys.

The advantage of using ReGenesees in the estimation process of the service turnover growth rate is that it provides the standard error related to the estimates of the totals (Chianella et al. 2013).

In 2013 different calibration models were tested (Bacchini et al. 2013). A comparison was made by integrating the annual turnover with other known information on the population. A list of different combinations of tested constraints is reported in the sequel. Annual turnover; annual turnover with employee number; annual turnover with company number; annual turnover with company and employee number. The analysis was conducted on a 3-year time interval (2010-2012). The range of the confidence intervals produced on the quarterly total estimates was evaluated together with the congruence at the

[7] See Zardetto D., 2015

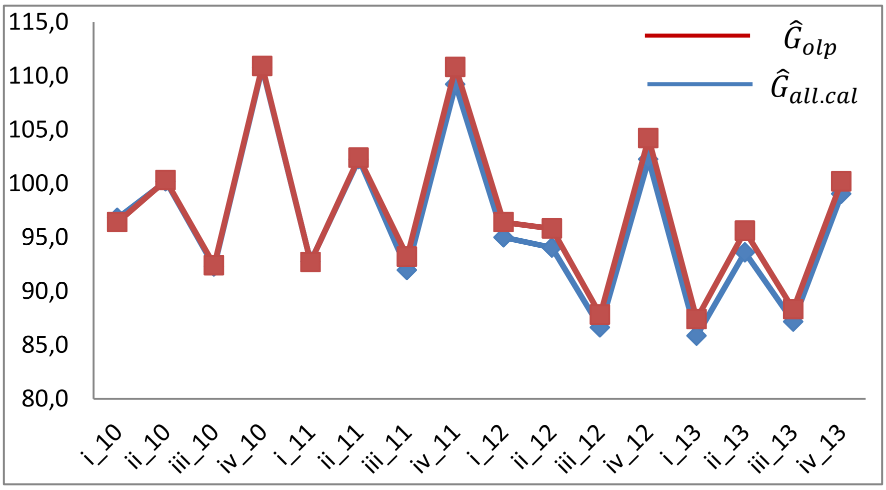
domain level between the totals annual estimates (obtained by the sum of the quarterly estimates) of 2010 and 2011, with the annual turnover in ASIA referred to 2010 and 2011. The results of the different models were very similar in terms of estimate values and in terms of coefficients of variation. However a smaller variability of calibrated weights within the stratum was observed for the model that used only the variable of annual turnover. With this model a lower correction factor of the initial weights was also observed.

2.2.3 - What estimator for the estimation of change over time?

Recently, to standardize the estimation methodology for new and old sectors, a debate has been opened to decide whether to adopt the estimator for calibration on all respondents in the two reference periods or an estimator based on the ratio between the estimates only on the companies in overlapping.

At the beginning of 2014, during the annual updating process of the sample, an application to real data was carried out for a comparison with the old estimator $\hat{G}_{olp}$ (Chianella et al. 2015[8]). The index of the maintenance and repair of motor vehicles was recalculated through the new estimator $\hat{G}_{all.cal}$, starting from the old base year (2010=100). Estimates based on the new estimator were obtained by considering the new stratification referred to 2014, and the results between the two estimated series were very similar (figure 2.2).

Figure 2.2. Index of maintenance and repair of motor vehicles (2010=100). Comparison between two estimators



[8] "An estimator for the growth rates in short-term business statistics using calibration" Journal of Official Statistics – Anniversary Conference 2015, June 10-12. D. Chianella, B. Iaconelli, R. Iannaccone (Short Term Statistics Directorate) -Poster Session.

As a consequence, from 2014 the index of maintenance and repair of motor vehicles sector (452 according to Nace Rev. 2) is calculated with the estimator $\hat{G}_{all.cal}$.

However, no assessment was performed for the standard error related to the growth rate estimation. To decide which estimator has to be used, it is necessary to analyze their standard errors in order to define the confidence intervals of the estimates. It has been just hypothesized that the estimator $\hat{G}_{all,cal}$ is better than the estimator $\hat{G}_{olp}$, because it provides accurate estimates on the totals, thanks to the high correlation between the variable of interest and the auxiliary calibration variable. Although the estimator $\hat{G}_{all,cal}$ provides a good standard error for the total estimations on both quarters, this does not mean that it is better in terms of variance for the estimate of the change g. Since the estimator for the percentage growth rate is defined as follows:

$$\hat{g} = \left(\hat{G} - 1\right)100\,,$$

and its variance is given by:

$$Var(\hat{g}) = 100^2 Var(\hat{G})$$

we can refer to the variance of G or to the variance of g to study the behavior of the variance of the estimators just proposed. In the next paragraph will be developed in detail the variance for the estimators of G discussed in this paragraph ($\hat{G}_{olp}$and $\hat{G}_{all.cal}$) both at the stratum and domain level

2.3 - The variance for the estimators of the change G: use of the first-order Taylor approximation

The $\hat{G}_{olp}$ and $\hat{G}_{all.cal}$ estimators are non-linear functions of linear estimators $(\hat{Y}^1, \hat{Y}^2)$. To calculate their variance we can use a first-order Taylor approximation, by approximating $\hat{G} = f(\hat{Y}^1, \hat{Y}^2)$ at the point (Y^1, Y^2). We can write:

$$\hat{G} = f(Y^1, Y^2) + \sum_{i=1}^{2} d_i(Y^1, Y^2)\left(\hat{Y}^i - Y^i\right) + O\left(\frac{1}{n}\right),$$

where $d_i(Y^1, Y^2)$ are the partial derivatives with respect to Y^1 and Y^2:

$$d_i(Y^1, Y^2) = \left[\frac{\partial f(\hat{Y}^1, \hat{Y}^2)}{\partial \hat{Y}^i}\right]_{\hat{Y}^i = Y^i} = d_i.$$

By replacing $f(Y^1, Y^2)$ with G, and by considering that the sample has a large size, the approximation

$$\hat{G} \cong G + \sum_{i=1}^{2} d_i \left(\hat{Y}^i, Y^i\right) = \left(G - \sum_{i=1}^{2} d_i Y^i\right) + \sum_{i=1}^{2} d_i \widehat{Y^\iota}$$

is obtained. The variance of $\hat{G}$ then becomes:

$$Var\left(\hat{G}\right) = Var\left(\left(G - \sum_{i=1}^{2} d_i Y^i\right) + \sum_{i=1}^{2} d_i \widehat{Y^\iota}\right).$$

The first term $\left(G - \sum_{i=1}^{2} d_i Y^i\right)$ is a constant, therefore we can write:

$$Var\left(\hat{G}\right) = Var\left(\sum_{i=1}^{2} d_i \widehat{Y^\iota}\right).$$

2.3.1 – Variance of the estimators of G within the stratum

To estimate the variance of G at the stratum level, using the $\hat{G}_{h,olp}$ and $\hat{G}_{h,all.cal}$ estimators, it is necessary for both cases to linearize a ratio. In fact, $\hat{G}_{h,olp}$ and $\hat{G}_{h,all.cal}$ can be written

$$\hat{G}_h = f\left(\hat{Y}^1, \hat{Y}^2\right) = \frac{\hat{Y}^1}{\hat{Y}^2},$$

where for the $\hat{G}_{h,olp}$ estimator $\hat{Y}^1 = \hat{\hat{y}}^t_{h,r2}$ and $\hat{Y}^2 = \hat{\hat{y}}^{t-4}_{h,r2}$ while for the $\hat{G}_{h,all.cal}$ estimator $\hat{Y}^1 = \hat{Y}^t_{h,r_{23}}$ and $\hat{Y}^2 = \hat{Y}^{t-4}_{h,r_{12}}$. For $i = 1$ and $i = 2$ we obtain the following partial derivates:

$$d_1 = \left[\frac{\partial f(\hat{Y}^1, \hat{Y}^2)}{\partial \hat{Y}^1}\right]_{\hat{Y}^1 = Y^1} = \frac{1}{Y^2}$$

$$d_2 = \left[\frac{\partial f(\hat{Y}^1, \hat{Y}^2)}{\partial \hat{Y}^2}\right]_{\hat{Y}^2 = Y^2} = -\frac{Y^1}{(Y^2)^2}.$$

As a consequence:

$$Var(\hat{G}_h) = Var\left(\sum_{i=1}^{2} d_i \hat{Y^i}\right) = Var\left(\frac{1}{Y^2}\hat{Y}^1 - \frac{Y^1}{(Y^2)^2}\hat{Y}^2\right) =$$

$$= Var\left(\frac{1}{Y^2}\left(\hat{Y}^1 - \frac{Y^1}{Y^2}\hat{Y}^2\right)\right) =$$

$$= \frac{1}{(Y^2)^2}\left\{Var(\hat{Y}^1) + \left(\frac{Y^1}{Y^2}\right)^2 Var(\hat{Y}^2) - 2\frac{Y^1}{Y^2} Cov(\hat{Y}^1, \hat{Y}^2)\right\} =$$

$$= \frac{1}{(Y^2)^2}\{Var(\hat{Y}^1) + G^2 Var(\hat{Y}^2) - 2G Cov(\hat{Y}^1, \hat{Y}^2)\}.$$

Therefore the variances of $\hat{G}_{h,olp}$ and $\hat{G}_{h,all.cal}$ are:

$$Var(\hat{G}_{h,olp}) = \frac{1}{\left(\bar{Y}_{h,r2}^{t-4}\right)^2}\{Var(\hat{\hat{y}}_{h,r2}^{t}) + G_h^2 Var(\hat{\hat{y}}_{h,r2}^{t-4}) - 2G_h Cov(\hat{\hat{y}}_{h,r2}^{t}, \hat{\hat{y}}_{h,r2}^{t-4})\}$$

$$Var(\hat{G}_{h,all.cal}) = \frac{1}{\left(Y_{h,r_{12}}^{t-4}\right)^2}\{Var(\hat{Y}_{h,r_{23}}^{t}) + G_h^2 Var(\hat{Y}_{h,r_{12}}^{t-4}) - 2G_h Cov(\hat{Y}_{h,r_{23}}^{t}, \hat{Y}_{h,r_{12}}^{t-4})\}$$

2.3.2 – Variance of the estimators of G within the estimation domain

When we are interested in the estimation of the variance of G at the domain level, using $\hat{G}_{d,olp}$ and $\hat{G}_{d,all.cal}$ estimators, we have to consider that the two estimators have a different form. In fact, the $\hat{G}_{d,olp}$ estimator is a sum of ratios while the $\hat{G}_{d,all.cal}$ estimator is a ratio of sums. This implies different calculations in the approximation in the Taylor series.

1. When we use the estimator $\hat{G}_{d,olp}$, by defining $c = \frac{1}{I_d^{t-4}}$, $b_h = I_h^{t-4} w_h$, $Y_h^1 = \hat{\hat{y}}_{h,r2}^{t}$ and $Y_h^2 = \hat{\hat{y}}_{h,r2}^{t-4}$, we can write:

$$\hat{G}_{d,olp} = c\sum_{h=1}^{H} \frac{\hat{Y}_h^1}{\hat{Y}_h^2}\, b_h.$$

For $i = 1$ and $i = 2$, taking into account the stratification, we obtain:

$$d_1 = \left[\frac{\partial f(\hat{Y}_h^1, \hat{Y}_h^2)}{\partial \hat{Y}_h^1}\right]_{\hat{Y}^1 = Y^1} = c\sum_{h=1}^{H} \frac{b_h}{Y_h^2} \;= c\sum_{h=1}^{H} \frac{b_h}{\hat{y}_{h,r2}^{t-4}}$$

$$d_2 = \left[\frac{\partial f(\hat{Y}^1, \hat{Y}^2)}{\partial \hat{Y}^2}\right]_{\hat{Y}^2 = Y^2} = -c\sum_{h=1}^{H} b_h \frac{Y_h^1}{(Y_h^2)^2}\ .$$

As a consequence, the variance of $\hat{G}_{d,olp}$ becomes:

$$Var\big(\hat{G}_{d,olp}\,\big) = Var\left(\sum_{i=1}^{2} d_i \hat{Y}_h^i\right) = Var\big(d_1\hat{Y}_h^1 + d_2\hat{Y}_h^2\big) =$$

$$= Var\,\left(c\sum_{h=1}^{H} \frac{b_h}{Y_h^2}\hat{Y}_h^1 - c\sum_{h=1}^{H} b_h \frac{Y_h^1}{(Y_h^2)^2}\,\hat{Y}_h^2\right) =$$

$$= c^2 Var\left(\sum_{h=1}^{H}\left(\frac{b_h}{Y_h^2}\hat{Y}_h^1 - b_h \frac{Y_h^1}{(Y_h^2)^2}\,\hat{Y}_h^2\right)\right) =$$

$$= c^2 Var\,\left(\sum_{h=1}^{H} b_h\left(\frac{\hat{Y}_h^1}{Y_h^2} - \frac{Y_h^1}{(Y_h^2)^2}\,\hat{Y}_h^2\right)\right) =$$

$$c^2\sum_{h=1}^{H} b_h^2 Var\left(\frac{1}{Y_h^2}\left(\hat{Y}_h^1 - \frac{Y_h^1}{Y_h^2}\,\hat{Y}_h^2\right)\right) =$$

$$= c^2\sum_{h=1}^{H} \frac{b_h^2}{(Y_h^2)^2} Var\left(\hat{Y}_h^1 - \frac{Y_h^1}{Y_h^2}\,\hat{Y}_h^2\right) =$$

$$= c^2\sum_{h=1}^{H} \frac{b_h^2}{(Y_h^2)^2}\left\{Var\big(\hat{Y}_h^1\big) + \left(\frac{Y_h^1}{Y_h^2}\right)^2 Var\big(\hat{Y}_h^2\big) - 2\frac{Y_h^1}{Y_h^2} Cov(\hat{Y}_h^1, \hat{Y}_h^2)\right\} =$$

$$= \frac{1}{(I_d^{t-4})^2}\sum_{h=1}^{H} \frac{(I_h^{t-4} w_h)^2}{(Y_h^2)^2}\left\{Var\big(\hat{Y}_h^1\big) + \left(\frac{Y_h^1}{Y_h^2}\right)^2 Var\big(\hat{Y}_h^2\big) - 2\frac{Y_h^1}{Y_h^2} Cov(\hat{Y}_h^1, \hat{Y}_h^2)\right\} =$$

$$= \frac{1}{(I_d^{t-4})^2}\sum_{h=1}^{H}\left((I_h^{t-4} w_h)^2 \frac{1}{(\bar{Y}_h^{t-4})^2}\left\{Var\big(\hat{\bar{y}}_{h,r2}^{t}\big) + G_h^2 Var\big(\hat{\bar{y}}_{h,r2}^{t-4}\big) - 2G_h Cov(\hat{\bar{y}}_{h,r2}^{t}, \hat{\bar{y}}_{h,r2}^{t-4})\right\}\right)$$

$$= \frac{1}{(I_d^{t-4})^2}\sum_{h=1}^{H}\left((I_h^{t-4}w_h)^2 Var(\hat{G}_{h,olp})\right)$$

2. When we use the estimator $\hat{G}_{d,all.cal}$ the calculation is similar to the one for the $\hat{G}_{h,all.cal}$ estimator, because they are both ratios. Therefore we can consider:

$$\hat{G} = f\left(\hat{Y}^1, \hat{Y}^2\right) = \frac{\hat{Y}^1}{\hat{Y}^2}.$$

$$Var\left(\hat{G}\right) = \frac{1}{(Y^2)^2}\left\{Var\left(\hat{Y}^1\right) + G^2 Var\left(\hat{Y}^2\right) - 2G Cov\left(\hat{Y}^1, \hat{Y}^2\right)\right\}$$

Therefore setting $\hat{Y}^1 = \sum_{h=1}^{H}\hat{Y}_{h,r_{23}}^{t}$ and $\hat{Y}^2 = \sum_{h=1}^{H}\hat{Y}_{h,r_{12}}^{t-4}$ we obtain:

$$Var\left(\hat{G}_{d,all.cal}\right) =$$

$$= \frac{1}{\left(\sum_{h=1}^{H} Y_{h,r_{12}}^{t-4}\right)^2}\left\{Var\left(\sum_{h=1}^{H}\hat{Y}_{h,r_{23}}^{t}\right) + G_d^2 Var\left(\sum_{h=1}^{H}\hat{Y}_{h,r_{12}}^{t-4}\right) - 2G_d Cov\left(\sum_{h=1}^{H}\hat{Y}_{h,r_{23}}^{t}, \sum_{h=1}^{H}\hat{Y}_{h,r_{12}}^{t-4}\right)\right\} =$$

$$= \frac{1}{(Y_d^{t-4})^2}\left\{\sum_{h=1}^{H} Var\left(\hat{Y}_{h,r_{23}}^{t}\right) + G_d^2\sum_{h=1}^{H} Var\left(\hat{Y}_{h,r_{12}}^{t-4}\right) - 2G_d\sum_{h=1}^{H} Cov\left(\hat{Y}_{h,r_{23}}^{t}, \hat{Y}_{h,r_{12}}^{t-4}\right)\right\} =$$

$$\frac{1}{(Y_d^{t-4})^2}\sum_{h=1}^{H}\left\{Var\left(\hat{Y}_{h,r_{23}}^{t}\right) + G_d^2\, Var\left(\hat{Y}_{h,r_{12}}^{t-4}\right) - 2G_d Cov\left(\hat{Y}_{h,r_{23}}^{t}, \hat{Y}_{h,r_{12}}^{t-4}\right)\right\}$$

In the above formula attention has to be paid at the term G_d, since the change is measured at the domain level and not at the stratum level.

2.3.3 – The variance terms within the Taylor approximation

We must distinguish between the two types of estimators used for the calculation of G.

1. Where the $\widehat{\boldsymbol{G}}_{\boldsymbol{olp}}$ estimator is used we have to calculate the variance of the turnover mean estimators for each stratum $h \in H$:

$$Var(\hat{\bar{y}}_{h,r2}^{t}) = \left(\frac{1}{n_{h,r2}} - \frac{1}{N_h}\right) S_{y_h^t}^2$$

$$Var(\hat{\bar{y}}_{h,r2}^{t-4}) = \left(\frac{1}{n_{h,r2}} - \frac{1}{N_h}\right) S_{y_h^{t-4}}^2$$

where $n_{h,r2}$ is the number of the overlapping respondent units to both quarters (*t* and *t-4)* and N_h is the number of units in the population within the stratum *h*. In this case, since we only use the turnover data of the units in the sample s_{23} for the estimation of both quarters, N_h is calculated from the ASIA version used to create the sample s_{23}. S_{h,y_t}^2 and $S_{h,y_{t-4}}^2$ are the adjusted population variance of the turnover within the stratum h, at the occasion t and t-4, respectively. S_{h,y_t}^2 and $S_{h,y_{t-4}}^2$can be estimated from the sample observations:

$$\hat{S}_{y_h^t}^2 = \frac{1}{n_{h,r2} - 1} \sum_{i \in r_2} \left(y_{h,i}^t - \bar{y}_{h,r_2}^t\right)^2$$

$$\hat{S}_{y_h^{t-4}}^2 = \frac{1}{n_{h,r2} - 1} \sum_{i \in r_2} \left(y_{h,i}^{t-4} - \bar{y}_{h,r_2}^{t-4}\right)^2$$

2. Where the $\widehat{\boldsymbol{G}}_{\boldsymbol{all.cal}}$ estimator is used, the calculation of the variance of the totals $\hat{Y}_{r_{23}}^t$and $\hat{Y}_{r_{12}}^{t-4}$ is more complex, because a calibration estimator is used. Let us see the methodology behind the calculation of the total variance when the calibration estimator is used.

   An important result obtained in Deville and Sarndal (1992) indicates that in large-scale surveys, calibration estimators that use a generic distance function are asymptotically equivalent to the corresponding generalized regression estimators using Euclidean distance. Therefore, the estimation of the variance of all the calibration estimator can be approximated by estimating the variance of the corresponding regression estimators, for which it is possible to derive the explicit expression.

   The variance of the estimated total using the generalized regression estimators is equal to the variance of the residuals. By following the steps in Righi et al. (2005) we can define the formula of the variance of the estimated total. We assume that the population U is divided into H strata and that the probability of inclusion of the units in the sample is constant within the stratum *h* (*h* = *1, ..., H*), so that $\pi_h = n_h / N_h$. In this context, the regression estimator can be calculated as follows:

$$\hat{Y}_{GREG} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k=1}^{n_h} \frac{y_{hk}}{\pi_h} g_{hk}$$

where *k* is the generic unit belonging to stratum *h* and g_{hk} is a correction factor of the initial weight $1/\pi_k = N_h/n_h$, defined in the following way:

$$g_{hk} = 1 + (X_h - \hat{X}_{HT,h}) \left(\sum_{k=1}^{n_h} \frac{x_{hk}^2}{\pi_h c_{hk}} \right)^{-1} \frac{x_{hk}}{c_{hk}}$$

where:

$$\hat{X}_{HT,h} = \sum_{k=1}^{n_h} \frac{x_{hk}}{\pi_{hk}} = \frac{n_h}{N_h} \sum_{k=1}^{n_h} x_{hk}$$

X_h is the known total of the calibration auxiliary variable within stratum h and $\hat{X}_h$ is its Horvitz-Thompson estimator computed by using sample observations. x_{hk} is the calibration auxiliary variable associated with the company k and c_{hk} is a known constant, usually fixed to 1. As mentioned before, the calibration variable used to generate estimations in the services turnover survey is the annual turnover of the enterprises.

If the sample units are selected without replacement, the variance estimation is calculated as

$$Var(\hat{Y}_{GREG}) = Var(\hat{Z}) =$$

$$= \sum\nolimits_{h=1}^{H} Var(\hat{Z}_h) =$$

$$\sum\nolimits_{h=1}^{H} Var\left(\sum_{k=1}^{N_h} \hat{z}_{hk} \right) =$$

$$= \sum\nolimits_{h=1}^{H} Var\left(\sum_{k=1}^{n_h} \frac{\hat{z}_{hk}}{\pi_h} g_{hk} \right) =$$

$$= \sum_{h=1}^{H} n_h \left(1 - \frac{n_h}{N_h}\right) \frac{\sum_{k=1}^{n_h} \left(\frac{\hat{z}_{hk}}{\pi_h} g_{hk} - \bar{Z}_h\right)^2}{n_h - 1} =$$

$$= \sum_{h=1}^{H} \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) \frac{\sum_{k=1}^{n_h} \left(\hat{z}_{hk} g_{hk} - \bar{\bar{Z}}_h\right)^2}{n_h - 1} =$$

$$= \sum_{h=1}^{H} N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) \frac{\sum_{k=1}^{n_h} \left(\hat{z}_{hk} g_{hk} - \bar{\bar{Z}}_h\right)^2}{n_h - 1},$$

$\hat{z}_{hk}$ is the Woodruff transform (Woodruff, 1971) and matches the term of the residuals $\hat{e}_k$ of the generalized regression model, while $\bar{\bar{Z}}_h$ is the mean of the quantity $\hat{z}_{hk} g_{hk}$ within the stratum h. The expression of the residuals $\hat{e}_k = \hat{z}_k$ of the generalized regression model is:

$$\hat{\boldsymbol{z}}_{\boldsymbol{hk}} = \hat{\boldsymbol{e}}_{\boldsymbol{hk}} = y_{hk} - \hat{y}_{hk} = y_{hk} - \hat{\beta}_h x_{hk}$$

Assuming that all values of the population (U) are known, it is possible to estimate the vector of the regression coefficients β through the use the weighted least squares method. Using the standard theory, the best unbiased linear estimator is given by:

$$\tilde{\beta}_h = \left(\sum_{k \in U} \frac{x_{hk} x'_{hk}}{c_{hk}}\right)^{-1} \sum_{k \epsilon U} \frac{x_{hk} y_{hk}}{c_{hk}} = T_{h1}^{-1} T_{h2}$$

However, the values of the variables X and Y are not known for all units of the population. An asymptotically correct estimate of β can be obtained by estimating T_{h1} and T_{h2} using the Horvitz-Thompson estimator:

$$\hat{T}_{h1} = \sum_{k \in s_{23}} \frac{x_{hk} x'_{hk}}{c_{hk} \pi_h}$$

$$\hat{T}_{h2} = \sum_{k \epsilon s_{23}} \frac{x_{hk} y_{hk}}{c_{hk} \pi_h}$$

Therefore the estimate of β_h is given by:

$$\hat{\beta}_h = \left(\sum_{k\in s} \frac{x_{hk}x'_{hk}}{c_{hk}\pi_{hk}}\right)^{-1} \sum_{k\epsilon s} \frac{x_{hk}y_{hk}}{c_{hk}\pi_h}$$

Residuals can be calculated as follows:

$$\hat{e}_{hk} = y_{hk} - \hat{\beta}_{hk}x_{hk} = y_{hk} - \left(\sum_{k\in s} \frac{x_{hk}x'_{hk}}{c_{hk}\pi_h}\right)^{-1} \sum_{k\epsilon s} \frac{x_{hk}y_{hk}}{c_{hk}\pi_h} x_{hk}$$

Therefore, the variance for the total estimates $\hat{Y}^{t}_{r_{23}}$ and $\hat{Y}^{t-4}_{r_{12}}$ can be computed by applying the above formulas. To compute $\hat{Y}^{t}_{r_{23}}$, the terms of the summation within $\hat{\beta}_h$ have to vary in the set $k \in r_{23}$, while to calculate $\hat{Y}^{t-4}_{r_{12}}$ in the set $k \in r_{12}$

As mentioned in the previous paragraphs, the ReGenesees software returns the variance of the totals estimation, speeding up the process.

2.3.4 – The covariance term within the Taylor approximation

In this section, the covariance $\mathrm{Cov}(\hat{Y}^1, \hat{Y}^2)$ is computed. In the Chapter 1 we have seen that under the assumption of a fixed population, sample size and overlapping rate as well as of the same stratification over time, the results of Tam (1984) and Qualité and Tillé (2008) can be easily derived (Andersson, 2011). In fact, the covariance of the mean estimator between two occasions can be expressed as follows:

$$Cov(\bar{y}^t_h, \bar{y}^{t-4}_h) = \left(\frac{n^{t-4,t}_h}{n^{t-4}_h n^t_h} - \frac{1}{N_h}\right) S_{y^t_h y^{t-4}_h} = \left(\frac{o_h}{n^t_h} - \frac{1}{N_h}\right) S_{y^t_h y^{t-4}_h}$$

where $S_{y^t_h y^{t-4}_h}$ is the adjusted population covariance of the turnover within stratum *h* between the occasions *t* and *t-4*, and o_h is the overlap given by the ratio between the number of common respondent units in both quarters and the number of respondent units in the quarter *t-4*.

We distinguish between the two types of estimators used for the calculation of G:

1. If we consider the $\boldsymbol{\widehat{G}_{olp}}$ estimator, we have to compute the covariance between two mean estimators, $\bar{y}_{r2}^{t}$ and $\bar{y}_{r2}^{t-4}$. We consider only the common respondents between the two occasions *t* and *t-4*. Therefore we have that $n_h^{t-4} = n_h^t = n_{h,r_2}$

$$Cov\left(\bar{y}_{h,r2}^{t}, \bar{y}_{h,r2}^{t-4}\right) = \left(\frac{1}{n_{h,r_2}} - \frac{1}{N_h}\right) S_{y_h^t y_h^{t-4}}$$

The adjusted population covariance $S_{y_h^t y_h^{t-4}}$ can be estimated from the sample observations:

$$\hat{S}_{y_h^t y_h^{t-4}} = \frac{1}{n_{h,r_2} - 1} \sum_{i \in r_2} \left(y_{h,i}^t - \bar{y}_{h,r2}^t\right)\left(y_{h,i}^{t-4} - \bar{y}_{h,r2}^{t-4}\right)$$

Therefore we can express the covariance estimate as:

$$\hat{C}ov\left(\bar{y}_{h,r2}^{t}, \bar{y}_{h,r2}^{t-4}\right) = \left(\frac{1}{n_{h,r_2}} - \frac{1}{N_h}\right) \hat{S}_{y_h^t y_h^{t-4}} =$$

$$= \left(\frac{1}{n_{h,r_2}} - \frac{1}{N_h}\right) \left\{ \frac{1}{n_{h,r_2} - 1} \sum_{i \in r_2} \left(y_{h,i}^t - \bar{y}_{h,r2}^t\right)\left(y_{h,i}^{t-4} - \bar{y}_{h,r2}^{t-4}\right) \right\}$$

Alternatively, the covariance between the two quarterly estimates can be calculated as the product of the auto-correlations between the estimates of the two quarters and the square root of the product of the related quarterly estimates (Ceccarelli et all. 2017). In formulas:

$$\hat{C}ov\left(\bar{y}_{h,r2}^{t}, \bar{y}_{h,r2}^{t-4}\right) = \hat{o}_h^{t-4,t}\ \hat{c}or\left(\bar{y}_{h,r2}^{t}, \bar{y}_{h,r2}^{t-4}\right) \sqrt{\hat{V}ar\left(\bar{y}_{h,r2}^{t}\right) \hat{V}ar\left(\bar{y}_{h,r2}^{t-4}\right)} =$$

$$= \hat{o}_h^{t-4,t}\ \hat{\rho}_h^{t-4,t} \sqrt{\left(\frac{1}{n_{h,r_2}} - \frac{1}{N}\right)^2 \hat{S}_{y_h^t}^2 \hat{S}_{y_h^{t-4}}^2} =$$

$$= \hat{\rho}_h^{t-4,t} \sqrt{\left(\frac{1}{n_{h,r_2}} - \frac{1}{N}\right)^2 \hat{S}_{y_h^t}^2 \hat{S}_{y_h^{t-4}}^2}$$

where $\hat{o}^{t-4,t}$ is the partial overlap between the estimates in the two quarters. In our case, since the G_{olp} estimator is used, $\hat{o}^{t-4,t} = 1$ because $n_h^{t-4} = n_h^t = n_{h,r_2}$ and there is complete overlap.

2. If we consider the $\boldsymbol{\widehat{G}_{all.cal}}$ estimator, we have to calculate the covariance between two totals estimators $\hat{Y}_{r_{23}}^{t}$ and $\hat{Y}_{r_{12}}^{t-4}$. We can use the formula:

$$Cov\left(\hat{Y}^{t}_{h,r_{23}}, \hat{Y}^{t-4}_{h,r_{12}}\right) = Cov\left(\hat{Z}^{t}_{h,r_{23}}, \hat{Z}^{t-4}_{h,r_{12}}\right) =$$
$$= N_h^2 \left(\frac{n^{t-4,t}_{h,r_2}}{n^{t-4}_{h,r_{12}} n^{t}_{h,r_{23}}} - \frac{1}{N_h} \right) S_{z^t_h z^{t-4}_h} =$$
$$= N_h^2 \left(\frac{o_{h,r_{12}}}{n^{t}_{h,r_{23}}} - \frac{1}{N_h} \right) S_{z^t_h z^{t-4}_h} \,,$$

where the covariance between the estimates is equal to the covariance between the residuals of the generalized regression model calculated at the occasion *t* on the set r_{23} and at the occasion *t-4* on the set r_{12}. $o_{h,r_{12}}$ is the overlapping of the respondent units between *t* and *t-4* with respect to the numer of respondent at the occasion *t-4* (set r_{12}).

The adjusted population covariance $S_{z^t_h z^{t-4}_h}$ can be estimated from the sample observations:

$$\hat{S}_{\tilde{z}^t_h \tilde{z}^{t-4}_h} = \frac{1}{n_{h,r_2} - 1} \sum_{i \in r_2} \left(z^t_{h,i} q^t_{h,i} - \bar{\tilde{z}}^t_{h,r2}\right)\left(z^{t-4}_{h,i} q^{t-4}_{h,i} - \bar{\tilde{z}}^{t-4}_{h,r2}\right),$$

where $\bar{\tilde{z}}^t_{h,r2}$ is the mean of $z^t_{h,i} q^t_{h,i}$ within the stratum *h*. Since the calculation of covariance only concerns the common observations to both quarters, the residuals are obtained from a regression model applied on the set r_2. In this case the correction factors of the initial weights corresponding to the unit "i" are the same in the two quarters ($q^t_{h,i} = q^{t-4}_{h,i}$). Therefore the estimator of $Cov\left(\hat{Y}^{t}_{h,r_{23}}, \hat{Y}^{t-4}_{h,r_{12}}\right)$ is:

$$\hat{C}ov\left(\hat{Y}^{t}_{h,r_{23}}, \hat{Y}^{t-4}_{h,r_{12}}\right) = N_h^2 \left(\frac{o_{h,r_{12}}}{n^{t}_{h,r_{23}}} - \frac{1}{N_h} \right) \frac{\sum_{i \in r_2} \left(z^t_{h,i} q^t_{h,i} - \bar{\tilde{z}}^t_{h,r2}\right)\left(z^{t-4}_{h,i} q^{t-4}_{h,i} - \bar{\tilde{z}}^{t-4}_{h,r2}\right)}{n_{h,r_2} - 1}$$

2.3.5 - The variance and covariance terms combined together

By combining the results of the variance and covariance terms together, we obtain the following further results.

a) The variance of the $\hat{G}_{olp}$ estimator within the stratum and within the estimation domains is equal to

$$Var\left(\hat{G}_{olp,h}\right) = Var\left(\frac{\hat{\bar{y}}_{h,r2}^{t}}{\hat{\bar{y}}_{h,r2}^{t-4}}\right) =$$

$$= \frac{1}{\left(\bar{Y}_{h,r_2}^{t-4}\right)^2}\left\{Var\left(\hat{\bar{y}}_{h,r2}^{t}\right) + G_h^2 Var\left(\hat{\bar{y}}_{h,r2}^{t-4}\right) - 2G_h cov\left(\hat{\bar{y}}_{h,r2}^{t}, \hat{\bar{y}}_{h,r2}^{t-4}\right)\right\} =$$

$$= \frac{1}{\left(\bar{Y}_{h,r_2}^{t-4}\right)^2}\left\{\left(\frac{1}{n_{h,r2}} - \frac{1}{N_h}\right)S_{y_{h,r2}^{t}}^2 + G_h^2\left(\frac{1}{n_{h,r2}} - \frac{1}{N_h}\right)S_{y_{h,r2}^{t-4}}^2 - 2G_h\left(\frac{1}{n_{h,r_2}} - \frac{1}{N_h}\right)S_{y_{h,r_2}^{t} y_{h,r_2}^{t-4}}\right\}$$

and its estimator is given by:

$$\hat{V}ar\left(\hat{G}_{h,olp}\right) = \hat{V}ar\left(\frac{\hat{\bar{y}}_{h,r2}^{t}}{\hat{\bar{y}}_{h,r2}^{t-4}}\right) =$$

$$= \frac{1}{\left(\hat{\bar{y}}_{h,r_2}^{t-4}\right)^2}\left\{\left(\frac{1}{n_{h,r2}} - \frac{1}{N_h}\right)\hat{S}_{y_{h,r2}^{t}}^2 + \hat{G}_{h,olp}^2\left(\frac{1}{n_{h,r2}} - \frac{1}{N_h}\right)S_{y_{h,r2}^{t-4}}^2 \right.$$
$$\left. - 2\hat{G}_{h,olp}\left(\frac{1}{n_{h,r_2}} - \frac{1}{N_h}\right)\hat{S}_{y_{h,r_2}^{t} y_{h,r_2}^{t-4}}\right\} =$$

$$= \frac{1}{\left(\bar{y}_{h,r_2}^{t-4}\right)^2}\left\{\left(\frac{1}{n_{h,r2}} - \frac{1}{N_h}\right)\left(\hat{S}_{y_{h,r2}^{t}}^2 + \hat{G}_{h,olp}^2\,\hat{S}_{y_{h,r2}^{t-4}}^2 - 2\hat{G}_{h,olp}\hat{S}_{y_{h,r_2}^{t} y_{h,r_2}^{t-4}}\right)\right\} =$$

$$= \frac{1}{\left(\bar{y}_{h,r_2}^{t-4}\right)^2}\left\{\left(\frac{1}{n_{h,r2}} - \frac{1}{N_h}\right)\hat{S}_{y_{h,r2}^{t} - \hat{G}_{h,olp} y_{h,r_2}^{t-4}}^2\right\},$$

where $\hat{G}_{h,olp} = \frac{\hat{\bar{y}}_{r2}^{t}}{\hat{\bar{y}}_{r2}^{t-4}} = \frac{\sum_{h=1}^{H}\hat{\bar{y}}_{h,r2}^{t}}{\sum_{h=1}^{H}\hat{\bar{y}}_{h,r2}^{t-4}}$. Once calculated the variance of the $\hat{G}_{olp}$estimator within the generic stratum *h,* the calculation of the variance estimate within the domain *d* is quite simple. It is equal to

$$\hat{V}ar\left(\hat{G}_{d,olp}\ \right) = \frac{1}{(I_d^{t-4})^2}\sum_{h=1}^{H}\left((I_h^{t-4}w_h)^2\hat{V}ar(\hat{G}_{h,olp})\right).$$

b) The variance of the $\hat{G}_{all.cal}$ estimator within the strata and the estimation domains is:

$$
\begin{aligned}
Var\left(\hat{G}_{h,all.cal}\right) &= Var\left(\frac{\hat{Y}^{t}_{h,r_{23}}}{\hat{Y}^{t-4}_{h,r_{12}}}\right) \\
&= \frac{1}{(Y_h^{t-4})^2}\left\{Var\left(\hat{Y}^{t}_{h,r_{23}}\right) + G_h^2 Var\left(\hat{Y}^{t-4}_{h,r_{12}}\right) - 2G_h cov\left(\hat{Y}^{t}_{h,r_{23}}, \hat{Y}^{t-4}_{h,r_{12}}\right)\right\} = \\
&= \frac{1}{(Y_h^{t-4})^2}\left\{Var\left(\hat{Z}^{t}_{h,r_{23}}\right) + G_h^2 Var\left(\hat{Z}^{t-4}_{h,r_{12}}\right) - 2G_h Cov\left(\hat{Z}^{t}_{h,r_{23}}, \hat{Z}^{t-4}_{h,r_{12}}\right)\right\},
\end{aligned}
$$

and its estimator is given by:

$$
\begin{aligned}
\hat{V}ar\left(\hat{G}_{h,all.cal}\right) &= \\
&= \frac{1}{\left(\hat{Y}_h^{t-4}\right)^2}\left\{N_h^2\left(\frac{1}{n_{h,r_{23}}} - \frac{1}{N_h}\right)\frac{\sum_{k\epsilon r_{23}}(\hat{z}^{t}_{hk}g_{hk} - \bar{Z}_h)^2}{n_{h,r_{23}} - 1}\right. \\
&+ \hat{G}^2_{h,all.cal}N_h^2\left(\frac{1}{n_{h,r_{12}}} - \frac{1}{N_h}\right)\frac{\sum_{k\epsilon r_{12}}(\hat{z}^{t}_{hk}g_{hk} - \bar{Z}_h)^2}{n_{h,r_{12}} - 1} \\
&\left. -2\hat{G}_{h,all.cal}N_h^2\left(\frac{o_{h,r_{12}}}{n_{h,r_{23}}} - \frac{1}{N_h}\right)\frac{\sum_{i\in r_2}\left(z^{t}_{h,i}q^{t}_{h,i} - \bar{\bar{z}}^{t}_{h,r2}\right)\left(y^{t-4}_{h,i}q^{t-4}_{h,i} - \bar{\bar{z}}^{t-4}_{h,r2}\right)}{n_{h,r_2} - 1}\right\}.
\end{aligned}
$$

As already mentioned, the relationships

$$
\hat{z}^{t-4}_{hk} = y^{t-4}_{hk} - \hat{y}^{t-4}_{hk} = y^{t-4}_{hk} - \hat{\beta}_{hk}x_{hk} = y^{t-4}_{hk} - \left(\sum_{k\in r_2}\frac{x^2_{hk}}{c_{hk}\pi_h}\right)^{-1}\sum_{k\epsilon r_2}\frac{x_{hk}y^{t-4}_{h,k}}{c_{hk}\pi_h}x_{hk}
$$

$$
\hat{z}^{t}_{hk} = y^{t}_{hk} - \hat{y}^{t}_{hk} = y^{t}_{hk} - \hat{\beta}_{hk}x_{hk} = y^{t}_{hk} - \left(\sum_{k\in r_2}\frac{x^2_{hk}}{c_{hk}\pi_h}\right)^{-1}\sum_{k\epsilon r_2}\frac{x_{hk}y^{t}_{h,k}}{c_{hk}\pi_h}x_{hk}
$$

$$
g_{hk} = 1 + (X_h - \hat{X}_{HT,h})\left(\sum_{k=1}^{n_h}\frac{x^2_{hk}}{\pi_h c_{hk}}\right)^{-1}\frac{x_{hk}}{c_{hk}}
$$

hold true. The term x_{hk} is the value of the calibration variable (annual turnover) for the unit k in the stratum h, and derives from the latest available Asia. X_h is the known total within the stratum h (the total annual turnover calculated from Asia).

To simplify formulas, we assume that within each stratum h the set of respondents is the same in both quarters ($n_{h,r_{12}} = n_{h,r_{23}} = n_h$). We define also the sample variances and the sample covariance of $\tilde{z}_h = \hat{z}_h^t g_h$ within the stratum h:

$$\hat{S}^2_{\tilde{z}_h^t, r_{23}} = \frac{\sum_{k\epsilon r_{23}} (\hat{z}_{hk}^t g_{hk} - \bar{Z}_h)^2}{n_{h,r_{23}} - 1}$$

$$\hat{S}^2_{\tilde{z}_h^{t-4}, r_{12}} = \frac{\sum_{k\epsilon r_{12}} (\hat{z}_{hk}^{t-4} g_{hk} - \bar{Z}_h)^2}{n_{h,r_{12}} - 1}$$

Therefore, we can write:

$$\hat{V}ar(\hat{G}_{h,all.cal}) = Var\left(\frac{\hat{Y}_{h,r_{23}}^t}{\hat{Y}_{h,r_{12}}^{t-4}}\right) =$$

$$= \frac{1}{\left(\hat{Y}_h^{t-4}\right)^2}\left\{N_h^2\left(\frac{1}{n_h} - \frac{1}{N_h}\right)\hat{S}^2_{\tilde{z}^t_{h,r_{23}}} + \ G^2_{h,all.cal} N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right)\hat{S}^2_{\tilde{z}^{t-4}_{h,r_{12}}} \right.$$
$$\left. - 2G_{h,all.cal} N_h^2\left(\frac{o_h}{n_h} - \frac{1}{N_h}\right)\hat{S}_{\tilde{z}_h^t \tilde{z}_h^{t-4}}\right\} =$$

$$= \frac{1}{\left(\hat{Y}_{h,r_{12}}^{t-4}\right)^2}\left\{N_h^2\left(\frac{1}{n_h} - \frac{1}{N_h}\right)\left(\hat{S}^2_{\tilde{z}^t_{h,r_{23}}} + \hat{G}^2_{h,all.cal}\hat{S}^2_{\tilde{z}^{t-4}_{h,r_{12}}}\right) - 2\hat{G}_{h,all.cal} N_h^2\left(\frac{o_h}{n_h} - \frac{1}{N_h}\right)\hat{S}_{\tilde{z}_h^t \tilde{z}_h^{t-4}}\right\}$$
$$= \frac{1}{\left(\hat{Y}_{h,r_{12}}^{t-4}\right)^2}\left\{N_h^2\left[\left(\frac{1}{n_h} - \frac{1}{N_h}\right)\hat{S}^2_{\tilde{z}^t_{h,r_{23}} - \hat{G}_{all,cal}\tilde{z}^{t-4}_{h,r_{12}}}\right]\right\}$$

and the estimator of the variance of $\hat{G}_{all.cal}$ within the estimation domain becomes:

$$\hat{V}ar(\hat{G}_{d,all.cal}) =$$

$$= \frac{1}{\left(\hat{Y}_d^{t-4}\right)^2}\left\{\sum_{h=1}^{H} Var\left(\hat{Y}_{h,r_{23}}^t\right) + \ \hat{G}^2_{d,all.cal}\sum_{h=1}^{H} Var\left(\hat{Y}_{h,r_{12}}^{t-4}\right)\right.$$
$$\left. - 2\hat{G}_{d,all.cal}\sum_{h=1}^{H} Cov(\hat{Y}_{h,r_{23}}^t, \hat{Y}_{h,r_{12}}^{t-4})\right\} =$$

$$= \frac{1}{\left(\hat{Y}_d^{t-4}\right)^2}\left\{\sum_{h=1}^{H} N_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)\hat{S}^2_{\tilde{z}^t_{h,r_{23}}} + \hat{G}^2_{d,all.cal}\sum_{h=1}^{H} N_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)\hat{S}^2_{\tilde{z}^{t-4}_{h,r_{12}}} \right.$$
$$\left. - 2\hat{G}_{d,all.cal}\sum_{h=1}^{H} N_h^2\left(\frac{o_{h,r_{12}}}{n^{t-4}_{h,r_{23}}}-\frac{1}{N_h}\right)\hat{S}_{z_h^t z_h^{t-4}}\right\}.$$

2.4 - When the estimator based on all respondent units is a better choice?

2.4.1 - Estimation without calibration

In the Chapter 1 we have seen (Qualité and Tillé, 2008) the comparison between the estimators $\hat{d}_c$ and $\hat{d}$ for the estimate of the difference between two quantities over time ($d = y^t - y^{t-4}$), where the estimator $\hat{d}_c = \hat{d}_{olp}$ is the difference between the sample means calculated only on the overlap observations between the two occasions, while $\hat{d} = \hat{d}_{all}$ is the difference between the sample means calculated on all observations.

Assuming $n_1 = n_2 = n$ and $\hat{\mathrm{S}}_{\mathrm{y}_1}{}^2 = \hat{\mathrm{S}}_{\mathrm{y}_2}{}^2 = S^2$, the authors obtained:

$$Var(\hat{d}) - Var(\hat{d}_c) \approx \left(\frac{1}{\mathrm{n}} - \frac{1}{E(n_{r_2})}\right)2\hat{S}^2 - \left(\frac{\mathrm{o}}{\mathrm{n}} - \frac{1}{E(n_{r_2})}\right)2\rho\hat{S}^2$$
$$= \frac{1}{no}\,[o-1]2\hat{S}^2 \; - \frac{\rho}{n\,o}[o^2-1]2\hat{S}^2 \; =$$
$$= \frac{2\hat{S}^2}{n\,o}(1-o)[\rho(1+o)-1]\,,$$

where $o = \frac{n_{r_2}}{n}$ and ρ is the correlation between y^t and y^{t-4}. From the expression the authors found out that:

$$\begin{cases} Var(\hat{d}_{all}) - Var(\hat{d}_{olp}) = 0 & if\ \rho = \dfrac{1}{(1+o)} \\ Var(\hat{d}_{all}) - Var(\hat{d}_{olp}) > 0 & if\ \rho > \dfrac{1}{(1+o)} \\ Var(\hat{d}_{all}) - Var(\hat{d}_{olp}) < 0 & if\ \rho < \dfrac{1}{(1+o)} \end{cases}$$

and highlighting for the overlap:

$$\begin{cases} Var(\hat{d}_{all}) - Var(\hat{d}_{olp}) = 0 & if\ o = \frac{1}{\rho} - 1 \\ Var(\hat{d}_{all}) - Var(\hat{d}_{olp}) > 0 & if\ o > \frac{1}{\rho} - 1 \\ Var(\hat{d}_{all}) - Var(\hat{d}_{olp}) < 0 & if\ o < \frac{1}{\rho} - 1 \end{cases}$$

Therefore it is clear that the estimator that uses only the overlap between the two sample ($\hat{d}_{olp}$) is better than the estimator $\hat{d}_{all}$ when $o > \frac{1}{\rho} - 1$. As we can see from Table 2.2, when the correlation coefficient between the variables over time is high, it is better to use the estimator $\hat{d}_{olp}$ (that considers the data on the overlap units between the quarters), also with a low overlap rate (when the correlation between y^t and y^{t-4} is equal to 0.95, an overlap of 5% is sufficient). This is the case for the service turnover survey, where the correlation between the observed variable is usually high (>0.9 with respect to the previous quarter of the same quarter of the previous year).

Table 2.2 – Overlapping value over which the estimator $\hat{\boldsymbol{d}}_{olp}$ is better than the estimator $\hat{\boldsymbol{d}}_{all}$. Analysis for different correlation values between $\mathbf{y^t}$ and $\mathbf{y^{t-4}}$

$\rho(y^t, y^{t-4})$	overlapping
0.5	1
0.6	0.67
0.7	0.43
0.8	0.25
0.9	0.11
0.95	0.05

Knottnerus (2012) in his analysis considers the estimators of the growth rate $g = \frac{y^t - y^{t-4}}{y^{t-4}}$, based on the estimated total at both occasions ($\hat{Y}^t$ and $\hat{Y}^{t-4}$), without using calibration. We indicate the estimator based on all respondent units in both quarters with $\hat{G}_{all}$, to distinguish it from the estimator $\hat{G}_{all,cal}$, where also the calibration is used. We also indicate with $\hat{G}_{olp}$ the estimator of the total based on the overlap units between the two occasion, as in the previous sections.

In case of a simple random sample without replacement and assuming no stratification to simplify the formulas, that the estimator $\hat{G}_{all}$ is defined as

$$\hat{G}_{all} = \frac{\hat{Y}^{t}_{r_{23}}}{\hat{Y}^{t-4}_{r_{12}}} = \frac{N\sum_{j\in r_{23}}\frac{y_j^t}{n_{23}}}{N\sum_{i\in r_{12}}\frac{y_i^{t-4}}{n_{12}}} = \frac{\bar{y}_{r_{23}}}{\bar{y}_{r_{12}}}.$$

Furthermore, assuming also $n_{12} = n_{23} = n$ its variance is:

$$Var\left(\hat{G}_{all}\right) = \frac{1}{\left(\bar{Y}^{t-4}_{r_{12}}\right)^2}\left\{Var\left(\bar{y}_{r_{23}}\right) + G^2 Var\left(\bar{y}_{r_{12}}\right) - 2G\, cov\left(\bar{y}_{r_{23}}, \bar{y}_{r_{12}}\right)\right\}$$
$$= \frac{1}{\left(\bar{Y}^{t-4}_{r_{12}}\right)^2}\left\{\left(\frac{1}{n} - \frac{1}{N}\right)\left(S^2_{y_t} + G^2 S^2_{y_{t-4}}\right) - 2G^2\left(\frac{o}{n} - \frac{1}{N}\right)S_{y_t y_{t-4}}\right\},$$

where $S^2_{y_t}$ and $S^2_{y_{t-4}}$are the adjusted population variance of the turnover at the occasion t and t-4 respectively, $S_{y_t y_{t-4}}$is the adjusted population covariance of the turnover between the two occasions and $\bar{Y}^{t-4}_{r_{12}}$ is the population mean of the turnover at the occasion $t-4$.

Knottnerus compares $\mathrm{Var}\left(\hat{\mathrm{G}}_{\mathrm{all}}\right)$ with $V\mathrm{ar}\left(\hat{\mathrm{G}}_{\mathrm{olp}}\right)$ without assuming $\mathrm{S_{y_1}}^2 = \mathrm{S_{y_2}}^2 = S^2$. He finds the overlapping value (o) for which $Var\left(\hat{G}_{olp}\right) = Var\left(\hat{G}_{all}\right)$. Above this value, the estimator $\hat{G}_{olp}$ performs better than the estimator $\hat{G}_{all}$, because:

$$Var\left(\hat{G}_{olp}\right) - Var\left(\hat{G}_{all}\right) =$$
$$= \frac{1}{\left(\bar{Y}^{t-4}_{r_2}\right)^2}\left\{\left(\frac{1}{n_{r2}} - \frac{1}{N}\right)\left(S^2_{y_t} + G^2 S^2_{y_{t-4}} - 2G S_{y_t y_{t-4}}\right)\right\}$$
$$- \frac{1}{\left(\bar{Y}^{t-4}_{r_{12}}\right)^2}\left\{\left(\frac{1}{n} - \frac{1}{N}\right)\left(S^2_{y_t} + G^2 S^2_{y_{t-4}}\right) - 2G\left(\frac{o}{n} - \frac{1}{N}\right)S_{y_t y_{t-4}}\right\}$$

$$= \frac{1}{\left(\bar{Y}^{t-4}_{r_{12}}\right)^2}\left\{\left(\frac{1}{n_{r_2}} - \frac{1}{n}\right)\left(S^2_{y_t} + G^2 S^2_{y_{t-4}}\right) - 2G\left(\frac{1}{n_{r_2}} - \frac{o}{n}\right)S_{y_t y_{t-4}}\right\}$$

$$= \frac{1}{n_{r2}\left(\bar{Y}^{t-4}_{r_{12}}\right)^2}\left\{(1-o)\left(S^2_{y_t} + G^2 S^2_{y_{t-4}}\right) - 2G\,(1-o^2)S_{y_t y_{t-4}}\right\}$$

$$= \frac{1-o}{n_{r2}\left(\bar{Y}^{t-4}_{r_{12}}\right)^2}\left(S^2_{y_t - G y_{t-4}} - 2GoS_{y_t y_{t-4}}\right).$$

Therefore $Var(\hat{G}_{olp}) = Var(\hat{G}_{all})$ when:

$$o = \frac{S^2_{y_t - Gy_{t-4}}}{2GS_{y_t y_{t-4}}}$$

and $Var(\hat{G}_{olp}) > Var(\hat{G}_{all})$ when $S_{y_t y_{t-4}} < 0$ or:

$$o < \frac{S^2_{y_t - Gy_{t-4}}}{2GS_{y_t y_{t-4}}}$$

provided that $S_{y_t y_{t-4}} > 0$.

2.4.2 - Estimation with calibration

When we use calibration, the procedure is the same used by Knottnerus, but the calculation must be made on the residuals of the generalized regression model. Using stratification we have:

$$Var(\hat{G}_{olp.cal}) - Var(\hat{G}_{all.cal}) =$$

$$= \frac{1}{(Y^{t-4})^2}\left\{\sum_{h=1}^{H} N_h^2 \left[\left(\frac{1}{n_{h,r2}} - \frac{1}{N_h}\right)\left(S^2_{\tilde{z}_h^t} + G^2 S^2_{\tilde{z}_h^{t-4}} - 2GS_{z_h^t z_h^{t-4}}\right)\right]\right\}$$

$$- \frac{1}{(Y^{t-4})^2}\left\{\sum_{h=1}^{H} N_h^2 \left[\left(\frac{1}{n_h} - \frac{1}{N_h}\right)\left(S^2_{\tilde{z}_h^t} + G^2 S^2_{\tilde{z}_h^{t-4}}\right) - 2G\left(\frac{o_h}{n_h} - \frac{1}{N_h}\right) S_{z_h^t z_h^{t-4}}\right]\right\}$$

$$= \frac{1}{(Y^{t-4})^2}\left\{\sum_{h=1}^{H} N_h^2 \left[\left(\frac{1}{n_{h,r2}} - \frac{1}{n_h}\right)\left(S^2_{\tilde{z}_h^t} + G^2 S^2_{\tilde{z}_h^{t-4}}\right) - 2G\left(\frac{1}{n_{h,r2}} - \frac{o_h}{n_h}\right) S_{\tilde{z}_h^{t-4}, \tilde{z}_h^t}\right]\right\} =$$

$$= \frac{1}{(Y^{t-4})^2}\left\{\sum_{h=1}^{H} \frac{N_h^2}{n_{h,r2}}\left[(1 - o_h)\left(S^2_{\tilde{z}_h^t} + G^2 S^2_{\tilde{z}_h^{t-4}}\right) - 2G(1 - o_h^2) S_{\tilde{z}_h^{t-4}, \tilde{z}_h^t}\right]\right\} =$$

$$= \frac{1}{(Y^{t-4})^2}\left\{\sum_{h=1}^{H} \frac{N_h^2(1 - o_h)}{n_{h,r2}}\left[S^2_{\tilde{z}_h^t - G\tilde{z}_h^{t-4}} - 2Go_h S_{\tilde{z}_h^{t-4}, \tilde{z}_h^t}\right]\right\}.$$

Therefore, sufficient condition for which $Var\left(\hat{G}_{olp.cal}\right) > Var\left(\hat{G}_{all.cal}\right)$ is that $S_{\tilde{z}_h^{t-4},\tilde{z}_h^t} < 0$ or:

$$o_h < \frac{S^2_{\tilde{z}_h^t - G\tilde{z}_h^{t-4}}}{2GS_{\tilde{z}_h^{t-4},\tilde{z}_h^t}} \qquad \forall h \epsilon H$$

provided that $\mathrm{S}_{\tilde{\mathrm{z}}_{\mathrm{h}}^{\mathrm{t-4}},\tilde{\mathrm{z}}_{\mathrm{h}}^{\mathrm{t}}} > 0$ within each stratum h.

In the next chapter, within a simulation study, we will calculate the thresholds overlap to understand when it is better to use all observations or only the overlapping observations. The test will be conducted with and without calibration, and will be repeated for different values of $\rho(y^t, y^{t-4})$ and for different correlation values between the variable of interest and the calibration variable.

CHAPTER 3

Simulation study

3.1 - Simulation in the case of non-stratified population

3.1.1 - Aim of the simulation

In the previous chapter four estimators were presented for the estimate of the year-over-year growth rate of the turnover:

$$g = \left(\frac{Y^t}{Y^{t-4}} - 1\right) * 100 = (G - 1) * 100.$$

where Y^t is the variable relative to the turnover at the quarter t and Y^{t-4} is the variable concerning the turnover at the quarter t-4. The four estimators described are summarized in the below table:

Table 3.1 – Estimators used in the simulation

Estimator of G	All respondent units	Only overlapping respondent units
Ratio of sample means	$\hat{\mathbf{G}}_{\mathbf{all}}$	$\hat{\mathbf{G}}_{\mathbf{olp}}$
Ratio of totals through calibration	$\hat{\mathbf{G}}_{\mathbf{all.cal}}$	$\hat{\mathbf{G}}_{\mathbf{olp.cal}}$

where:

1. $\hat{G}_{olp}$ is based on the ratio of the sample means calculated by using turnover data on the overlapping respondent units (r_2) between the two quarters:

$$\hat{G}_{olp} = \frac{\bar{y}_{r_2}^{t}}{\bar{y}_{r_2}^{t-4}}$$

2. $\hat{G}_{all}$ is based on the ratio of the sample means calculated using turnover data on all respondent units over the two quarters:

$$\hat{G}_{all} = \frac{\bar{y}_{r_{23}}^{t}}{\bar{y}_{r_{12}}^{t-4}}$$

3. $\hat{G}_{olp.cal}$ is based on the ratio of the estimated total of the turnover for the quarter t and for the quarter t-4, calculated using turnover data on the overlapping respondent units between the two quarters and through calibration of the initial weights:

$$\hat{G}_{olp.call} = \frac{\hat{Y}_{r_2}^{t}}{\hat{Y}_{r_2}^{t-4}} = \frac{\sum_{j \in r_2} y_j^t w_j}{\sum_{i \in r_2} y_i^{t-4} w_i}$$

4. $\hat{G}_{all.cal}$ is based on the ratio of the estimated total of the turnover for the quarter t and the quarter $t-4$, calculated using turnover data on all respondent units over the two quarters and through calibration of the initial weights:

$$\hat{G}_{all.call} = \frac{\hat{Y}_{r_{23}}^{t}}{\hat{Y}_{r_{12}}^{t-4}} = \frac{\sum_{j \in r_{23}} y_j^t w_j}{\sum_{i \in r_{12}} y_i^{t-4} w_i},$$

   the calibrated weights (w_j and w_i) associated with the same unit on the two survey occasions of investigation (t and $t-4$) can be different due to the different non-response on the two occasions (the sets of respondent enterprises r_{12} and r_{23} are not the same).

A simulation study was conducted with the aim of analyzing the performance of these estimators. The bias, the standard deviation and the mean squared error have been analyzed through 1000 different samples extracted from the population and considering the following elements:

- Different values of the overlap (o) between the units responding at the occasion t and the units responding at the occasion t-4. In particular, the results have been analyzed by considering overlapping of 5%, 10%, 15%, 20%, 25%, 30%, 50%, 70%, 99%.

- Different values of the correlation between the variable of interest and the calibration variable. In particular, the results have been analyzed by considering correlation coefficient values $rho = 0, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95,1$.

- Different correlation values (0.97, 0.92 and 0.87) between the study variable on the two survey occasions $Y^t$and Y^{t-4} (as explained in Section 3.1.3).

3.1.2 - Main simulation steps

The main steps in the simulation study can be summarized in this way.

- A population of $N = 8360$ units has been generated with turnover possessing a lognormal distribution with parameters (mean and variance) able to reproduce the population observed in the sector of Accommodation, in the size class between 2 and 5 employees. The population generated represents the universe at the occasion t-4.

- The generation of the population data at the occasion t has been obtained assuming the following model:

$$Y_i^t = \beta Y_i^{t-4} + \varepsilon_i\, Y_i^{t-4}$$
$$\varepsilon \sim N(0, \sigma^2)$$

  The value of β has been fixed equal to 0.9. ε_i is a random variable with normal distribution. The increase of the variance σ^2 leads to a greater data variability at the occasion t, and to a lower correlation between the data at the occasions t and t-4.

- A calibration variable has been created according to the desired correlation with the interest variable Y^t. The created calibration variable has the same values for both occasions t and t-4. This make the simulation as similar as possible to the estimation process used for the estimation of the change in the service sector turnover in Istat. In fact, in this case, for both occasions the calibration variable coincides with the information available from the latest available Asia.

- The sample size is calculated from the population at the occasion t-4, by means of the Bethel algorithms implemented in Mauss-R (see Barcaroli et al, 2010). The planned coefficient of variation for the estimation of the total turnover has been fixed at 3%. The result is a sample size of $n = 417$ units.

- The sample at the occasion t-4 has been selected from the reference population. A random non-response of 30% of the units in the sample has been applied. Therefore, the size of the set of respondent units is equal to $\mathrm{n_r} = 292$.

- The sample at the occasion t consists in the union of a random subset of the respondent units at the occasion t-4, with size $\mathrm{n_c}$ depending on the desired overlapping o ($\mathrm{n_c} = \mathrm{on_r}$) and of a srswor of size $\mathrm{n_r} - \mathrm{n_c}$ from the population (excluding the units in the first subset). Therefore the size of respondent units is the same in both occasions.

- The estimates of the growth rate of the total turnover in the population between the two occasions are calculated using the four estimators above described. Beside, assuming normality and using Student's t-values, a confidence interval at 95% level is calculated for each estimate.

- The estimates are calculated on 1000 different samples selected from the reference populations. This allows the calculation of the bias, the standard deviation and the mean squared error for each estimator used.

3.1.3 - Simulation results

Three different simulations were performed, by fixing the variance parameter σ^2 in ε at the values 0.15, 0.25 and 0.35. The Y_i^{t-4} values are the same for the three different simulations, while the Y_i^t values depend on parameter σ^2 in ε. In Figure 3.1, 3.2 and 3.3 the graphs of the distribution of the Y values at the occasion t-4 (x-axis) and at the occasion t (y-axis) for the three different simulations are reported. The resulting correlation coefficients between $Y^t$and Y^{t-4} are respectively 0.97, 0.92 and 0.86, while the true values of the growth rate to be estimated resulting from the simulations are -10.0%, -10.5% and -10.2% respectively (Table 3.1).

The overlapping values for which the variance of the estimator based only on the overlapping units between both occasions (with or without calibration) is greater than the estimator that uses all available data in both occasions, have been calculated. As defined in the previous chapter, the calculation is performed according to the formula:

- $Var(\hat{G}_{olp}) > Var(\hat{G}_{all})$ when:

$$o < \frac{S^2_{Y^t - GY^{t-4}}}{2GS_{Y^{t-4},Y^t}}$$

provided $S_{Y^{t-4},Y^t} > 0$ and the calibration is not used.

- $Var(\hat{G}_{olp.cal}) > Var(\hat{G}_{all.cal})$ when:

$$o < \frac{S^2_{Z^t - GZ^{t-4}}}{2GS_{Z^{t-4},Z^t}}$$

provided $S_{Z^{t-4},Z^t} > 0$ and the calibration is used.

In the first case, the calculation takes into account the values of turnover in the population, while in the second case it takes into account the residuals of the generalized regression models. The simulation is performed for different overlapping rates, values of the correlation between the variable of interest and the calibration variable and different correlations between Y^t and Y^{t-4}.

Figure 3.1 – Plot of Y_i^t and Y_i^{t-4}. Simulation 1: ε ~ N(0, 0.15)

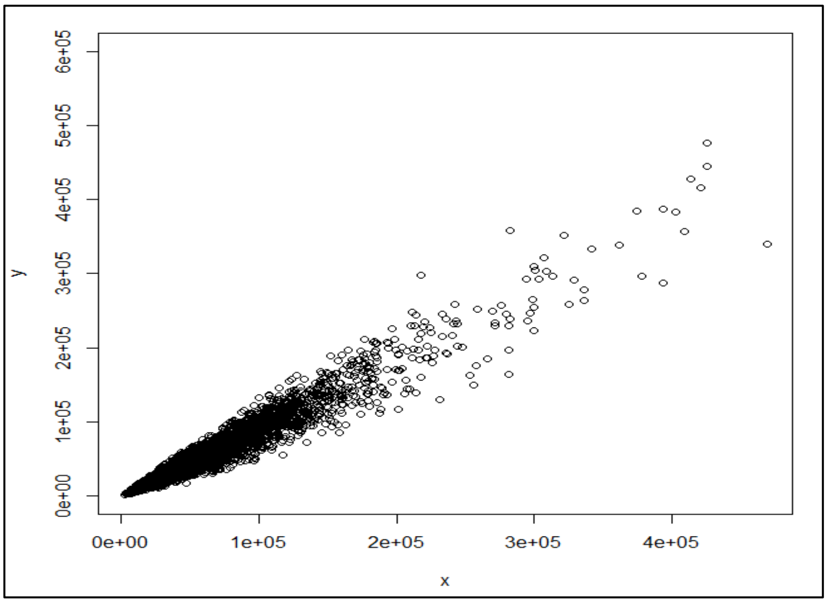


Figure 3.2 – Plot of Y_i^t and Y_i^{t-4}. Simulation 2: ε ~ N(0, 0.25)

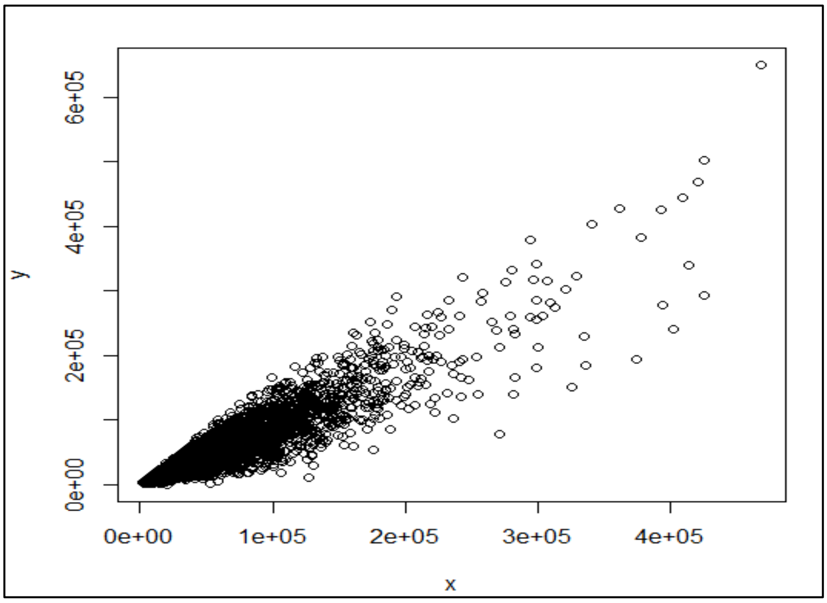


Figure 3.3 – Plot of Y_i^t and Y_i^{t-4}. Simulation 3: ε ~ N(0, 0.35)

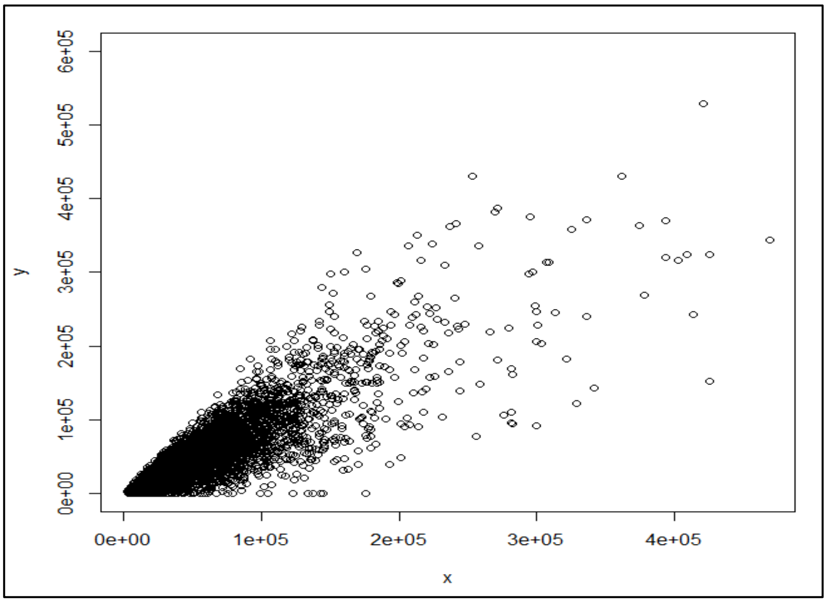


Table 3.1 – Correlation between the data over time and the growth rate in the population. Three different simulations

Simulation	$cor(Y^t, Y^{t-4})$	G	percentage growth rate g
1: $\varepsilon \sim N(0,0.15)$	0.97	0.900	-10.0
2: $\varepsilon \sim N(0,0.25)$	0.92	0.895	-10.5
3 $\varepsilon \sim N(0,0.35)$	0.86	0.898	-10.2

Table 3.2 shows the theoretical overlapping value (o) below which $Var(\hat{G}_{olp.cal}) > Var(\hat{G}_{all.cal})$. The following remarks are drawn.

- At the same rho values, when the variability of the data in the population at the occasion t increases (therefore decreases the correlation between Y_i^t and Y_i^{t-4}), the overlapping value (o) below which $Var(\hat{G}_{olp.cal}) > Var(\hat{G}_{all.cal})$ increases, as well. This is because the higher variability of Y^t will result in a higher variance of the residuals $S_{Z^t}^2$ (see Table 3.4). At the same time, the correlation between Y^{t-4} and the calibration variable decreases, because the calibration variable is created according to the desired correlation with the last available data x (Y^t). For this reason, the variance of the residuals $S_{Z^{t-4}}^2$ also increases (Table 3.3). Therefore the numerator of the o threshold value ($S^2_{Z^t-GZ^{t-4}}$) will become greater. On the other hand, the covariance between the residuals S_{Z^{t-4},Z^t} remains stable (see Table 3.5).

- When the correlation between the variable of interest and the calibration variable (rho) increases, then the overlapping value (o) below which $Var(\hat{G}_{olp.cal}) > Var(\hat{G}_{all.cal})$ will increase too. This is because the covariance between the residuals of the generalized regression model (S_{Z^{t-4},Z^t}) at the denominator of the o threshold value decreases (see Table 3.5). On the other hand, in the numerator the decrease in the covariance of the residuals counterbalances the decrease in the variances of the residuals (because the covariance is negative).

Table 3.2 – Theoretical overlapping values (o) below which $Var(\hat{G}_{olp.cal}) > Var(\hat{G}_{all.cal})$

rho	Simulation 1: $\varepsilon \sim N(0,0.15)$ $cor(Y^t, Y^{t-4}) = 0.97$	Simulation 2: $\varepsilon \sim N(0,0.25)$ $cor(Y^t, Y^{t-4}) = 0.92$	Simulation 3: $\varepsilon \sim N(0,0.35)$ $cor(Y^t, Y^{t-4}) = 0.86$
0	0.03	0.09	0.17
0.5	0.04	0.11	0.22
0.6	0.05	0.13	0.25
0.7	0.06	0.16	0.30
0.8	0.08	0.22	0.41
0.9	0.15	0.40	0.74
0.95	0.29	0.79	1
1	1	1	1

Table 3.5 shows the correlation and the covariance between the residuals of the generalized regression model for different rho values, calculated on population data. As we can see from the table, when rho increases, the correlation $cor(Z^t, Z^{t-4})$ and the covariance S_{Z^{t-4},Z^t} between the residuals of the regression models decreases. When rho is equal to 0, obviously $cor(Z^t, Z^{t-4}) = cor(Y^t, Y^{t-4})$. In the extreme case that rho = 1, the correlation and the covariance between the residuals of the models are equal to 0.

Table 3.3 – Variance of the residuals for Y^{t-4}

rho	Simulation 1: $\varepsilon \sim N(0,0.15)$	Simulation 2: $\varepsilon \sim N(0,0.25)$	Simulation 3: $\varepsilon \sim N(0,0.35)$
0	1,712,458,364	1,712,165,298	1,712,350,443
0.5	1,312,720,626	1,353,516,782	1,428,411,137
0.6	1,122,304,677	1,184,301,182	1,249,845,643
0.7	911,148,147	985,824,164	1,081,948,932
0.8	688,393,677	776,447,713	895,087,029
0.9	391,901,001	539,355,264	703,223,565
0.95	253,657,697	394,235,598	571,985,065
1	96,365,726	252,506,239	450,809,933

Figures 3.4-3.7 show the regression models for different rho values in both occasions (t and t-4) as well as the plot of the residuals of the models obtained in the second simulation (*ε ~ N(0, 0.25))*. The regressions are computed on population data. As it can be seen from the graphs, when rho increases, the covariance between the residuals decreases. It is perfectly clear that in the case rho = 1, the calibration variable is perfectly "aligned" to Y^t, while this does not happen for Y^{t-4}. In fact, for Y^{t-4} the residuals are always higher than those for Y^t.

Table 3.4 – Variance of the residuals for $\boldsymbol{Y^t}$

rho	Simulation 1: $\varepsilon \sim N(0{,}0.15)$	Simulation 2: $\varepsilon \sim N(0{,}0.25)$	Simulation 3: $\varepsilon \sim N(0{,}0.35)$
0	1,478,390,155	1,592,154,862	1,784,634,978
0.5	1,114,379,559	1,196,564,813	1,355,677,938
0.6	937,557,300	1,022,610,209	1,135,100,241
0.7	745,861,203	813,614,571	892,971,191
0.8	539,493,015	562,004,402	644,144,109
0.9	272,432,479	303,971,319	346,183,167
0.95	144,393,543	155,796,167	173,323,548
1	0	0	0

Table 3.5 – Correlation and Covariance of the residuals for $\boldsymbol{Y^{t-4}}$*and* $\boldsymbol{Y^t}$

	Simulation 1: $\varepsilon \sim N(0{,}0.15)$		Simulation 2: $\varepsilon \sim N(0{,}0.25)$		Simulation 3: $\varepsilon \sim N(0{,}0.35)$	
rho	$cor\ (Z^t, Z^{t-4})$	S_{Z^{t-4},Z^t}	$cor\ (Z^t, Z^{t-4})$	S_{Z^{t-4},Z^t}	$cor\ (Z^t, Z^{t-4})$	S_{Z^{t-4},Z^t}
0	0.97	1,545,721,412	0.92	1,524,497,200	0.86	1,500,463,050
0.5	0.96	1,164,252,656	0.90	1,147,800,874	0.83	1,151,467,642
0.6	0.96	980,753,386	0.89	976,160,357	0.80	952,363,642
0.7	0.95	779,560,614	0.86	772,481,286	0.76	750,725,615
0.8	0.93	565,152,010	0.82	542,668,332	0.70	535,019,354
0.9	0.87	283,751,417	0.73	295,323,943	0.60	295,647,544
0.95	0.79	150,705,087	0.60	148,597,686	0.46	144,924,850
1	0.01	0	0.03	0	0.00	0

Figure 3.4 – Regression models for Y^t and Y^{t-4} with the calibration variable and residuals plot (Z^{t-4}, Z^t). Case ε ~ N(0, 0.25) and rho=0

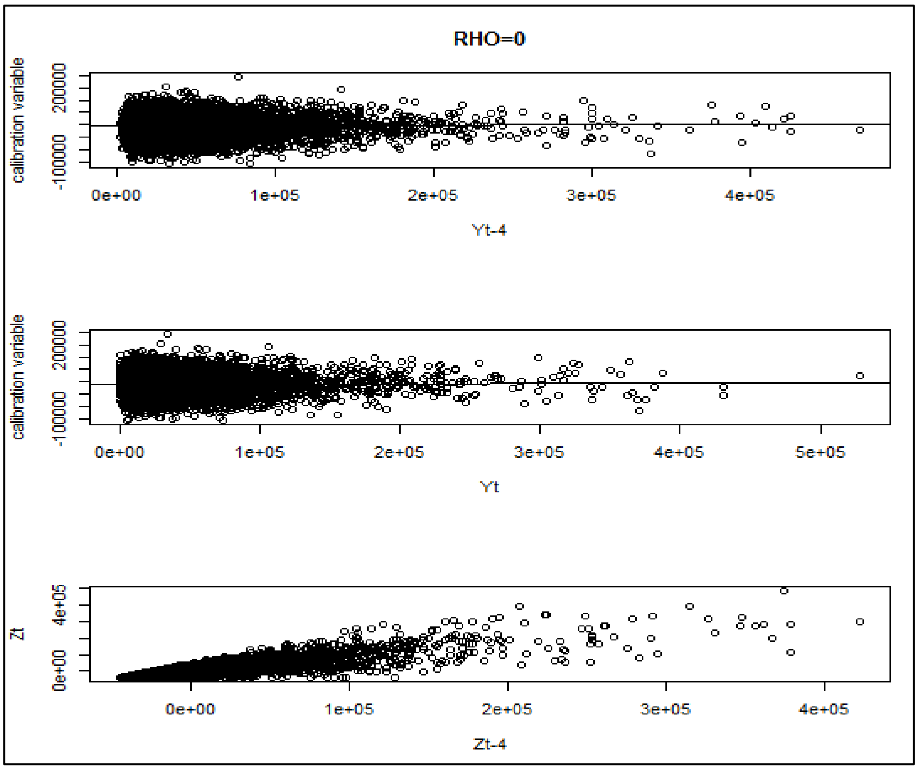


Figure 3.5 – Regression models for Y^t and Y^{t-4} with the calibration variable and residuals plot (Z^{t-4}, Z^t). Case ε ~ N(0, 0.25) and rho=0.8

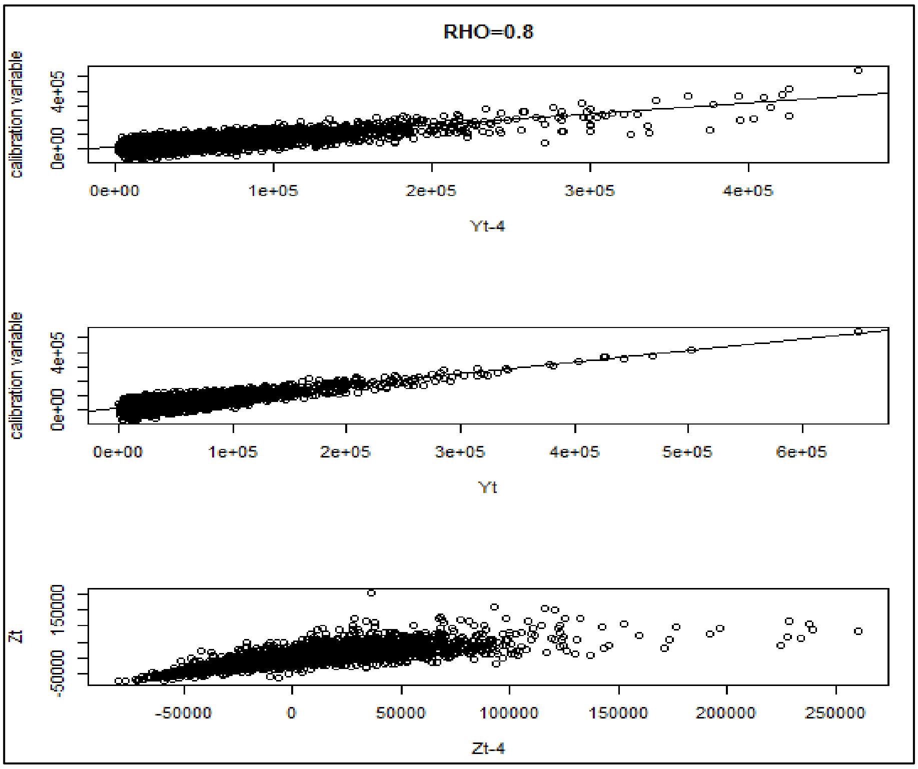


Figure 3.6 – Regression models for Y^t and Y^{t-4} with the calibration variable and residuals plot (Z^{t-4}, Z^t). Case ε ~ N(0, 0.25) and rho=0.95

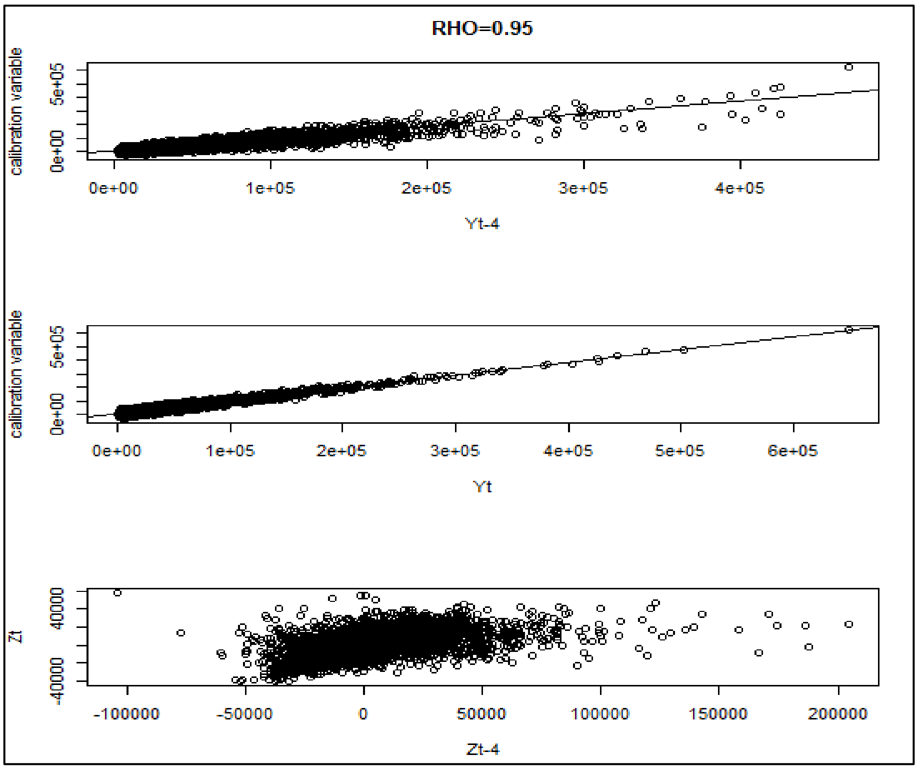
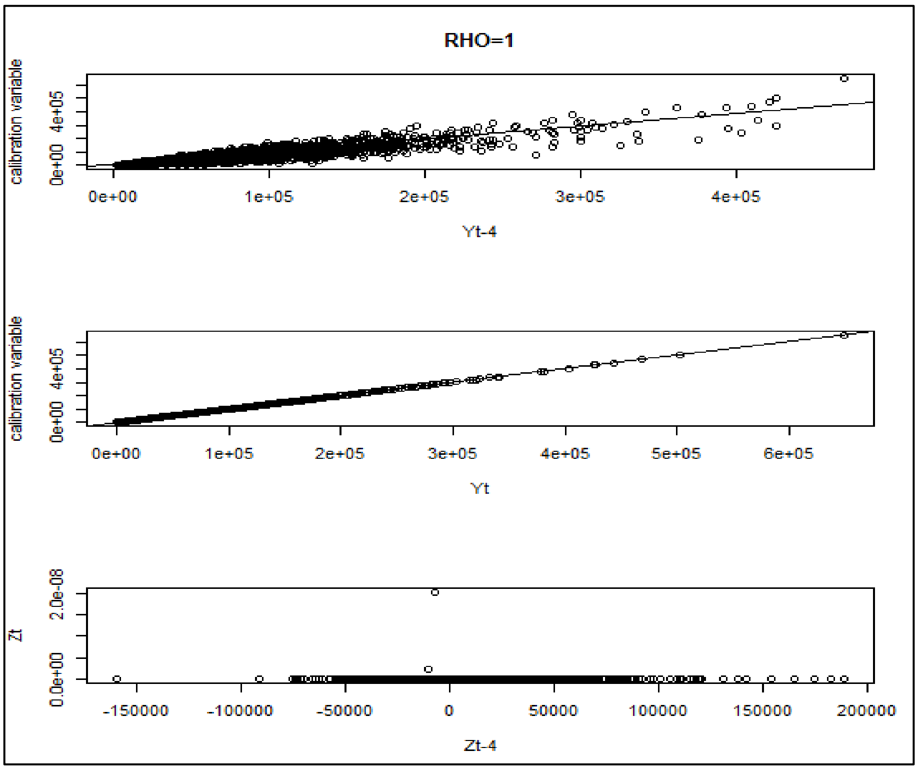


Figure 3.7 – Regression models for Y^t and Y^{t-4} with the calibration variable and residuals plot (Z^{t-4}, Z^t). Case ε ~ N(0, 0.25) and rho=1



Since we know the turnover values for each unit of the population, it is possible to compute the value of the standard deviation for each estimator of the growth rate. Since the size of the common respondents in both occasions is $n_c = on_r$, we have that:

1. When we do not use calibration, we have to compute the variance of the turnover data in the population, therefore we have to use the following formulas

$$Var(\hat{Y}_{all}) = N^2\left(\frac{1}{n_r} - \frac{1}{N}\right)S_Y^2 \ ,$$

$$Var(\hat{Y}_{olp}) = N^2\left(\frac{1}{n_c} - \frac{1}{N}\right)S_Y^2 \ ,$$

$$Cov(\hat{Y}_{all}^{t-4}, \hat{Y}_{all}^{t}) = \sum_{h=1}^{H} N^2\left(\frac{o}{n_r} - \frac{1}{N}\right)S_{Y^t,Y^{t-4}} \ ,$$

$$Cov(\hat{Y}_{olp}^{t-4}, \hat{Y}_{olp}^{t}) = \sum_{h=1}^{H} N^2\left(\frac{1}{n_c} - \frac{1}{N}\right)S_{Y^t,Y^{t-4}} \ ,$$

$$Var(\hat{G}_{all}) = \frac{1}{Y^{t-4}}\{Var(\hat{Y}_{all}^{t}) + G^2Var(\hat{Y}_{all}^{t-4}) - 2Gcov(\hat{Y}_{all}^{t-4}, \hat{Y}_{all}^{t})\} \ ,$$

$$Var(\hat{G}_{olp}) = \frac{1}{Y^{t-4}}\{Var(\hat{Y}_{olp}^{t}) + G^2Var(\hat{Y}_{olp}^{t-4}) - 2GCov(\hat{Y}_{olp}^{t-4}, \hat{Y}_{olp}^{t})\} \ .$$

2. When we use calibration, we have to compute the variance of the residuals of the regression model (S_Z^2) for each rho value. For each overlapping level of the respondent units between the two occasions (o), we have to compute:

$$Var(\hat{Z}_{all.cal}) = N^2\left(\frac{1}{n_r} - \frac{1}{N}\right)S_Z^2 \ ,$$

$$Var(\hat{Z}_{olp.cal}) = N^2\left(\frac{1}{n_c} - \frac{1}{N}\right)S_Z^2 \ ,$$

$$Cov(\hat{Z}_{all.cal}^{t-4}, \hat{Z}_{all.cal}^{t}) = \sum_{h=1}^{H} N^2\left(\frac{o}{n_r} - \frac{1}{N}\right)S_{Z^t,Z^{t-4}} \ ,$$

$$Cov\left(\hat{Z}_{olp.cal}^{t-4}, \hat{Z}_{olp.cal}^{t}\right) = \sum_{h=1}^{H} N^2 \left(\frac{1}{n_c} - \frac{1}{N}\right) S_{z^t, z^{t-4}},$$

Therefore, it is possible to compute the variance of the estimator $\hat{G}_{all.cal}$ and $\hat{G}_{olp.cal}$ through the expressions:

$$Var\left(\hat{G}_{all.cal}\right) = \frac{1}{Y^{t-4}} \left\{ Var\left(\hat{Z}_{all.cal}^{t}\right) + G^2 Var\left(\hat{Z}_{all.cal}^{t-4}\right) - 2Gcov\left(\hat{Z}_{all.cal}^{t-4}, \hat{Z}_{all.cal}^{t}\right) \right\},$$

$$Var\left(\hat{G}_{olp.cal}\right) = \frac{1}{Y^{t-4}} \left\{ Var\left(\hat{Z}_{olp.cal}^{t}\right) + G^2 Var\left(\hat{Z}_{olp.cal}^{t-4}\right) - 2GCov\left(\hat{Z}_{olp.cal}^{t-4}, \hat{Z}_{olp.cal}^{t}\right) \right\}.$$

Since $g = (G - 1) * 100$, for each estimator of g, we have that:

$$Var(\hat{g}) = 100^2 \, Var\left(\hat{G}\right).$$

Therefore from the variance of $\hat{G}$ it is possible to calculate the standard errors of $\hat{g}_{all.cal}$ and $\hat{g}_{olp.cal}$:

$$Se(\hat{g}) = 100\sqrt{Var\left(\hat{G}\right)}.$$

In Tables 3.7-3.9, the theoretical standard deviations of the above estimators are shown. The computed values are obtained from the calculation on the populations generated in the 3 simulation exercises. Instead, in Tables 3.10-3.12 the standard deviations of the 1000 sample estimates are shown. These values are calculated for different rho values and different overlapping between the respondent units at both occasions.

The standard deviations of the $\hat{G}_{all}$ and $\hat{G}_{olp}$ estimators are the same for each rho value because they do not need calibration. However, the tables show the values for each rho in order to simplify the comparison of the behavior of all estimators.

The variable *"o"* in the tables shows the theoretical overlapping values for which the estimator using only the overlapping respondent units between both occasions is greater than the estimator using all respondent unit values in both occasions. Its value is calculated according to the formulas described above.

The colored parts in the tables indicate the $Se(\hat{G}_{olp.cal}) > Se(\hat{G}_{all.cal})$ if we use the calibration estimators, or that $Se(\hat{G}_{olp}) > Se(\hat{G}_{all})$ if we do not use the calibration estimators.

As we can see from the results of the calculation of the standard deviations, when the overlap of the respondent units between the occasions increases, the standard deviation of all estimators decrease. This is in accordance with the theory in Chapter 1, because the variance of the change takes minimum value in the case of complete overlap (Kish, 1965, pp. 457-466).

Using calibration we obtain the best results, therefore we have that $Se(\hat{G}_{all.cal}) \leq Se(\hat{G}_{all})$ and that $Se(\hat{G}_{olp.cal}) \leq Se(\hat{G}_{olp})$ for each rho $\neq 0$ and for every overlap value. In particular, the greatest improvement is obtained when using the estimators based on all respondents ($\hat{G}_{all.cal}$ VS $\hat{G}_{all}$), while we observed only a limited improvement when using the estimators based on the overlap respondents ($\hat{G}_{olp.cal}$VS $\hat{G}_{olp}$). In this last case, the use of calibration leads to a smaller improvement because the calibration variable (X) is the same for both occasions (t and t-4). As consequence, since the initial weights π are the same for all units, if the variability of the correction factor (g_i) between the units is small, then the result obtained by the $\hat{G}_{olp.cal}$ estimator is similar to the one obtained by $\hat{G}_{olp}$:

$$\hat{G}_{olp.cal} = \frac{\sum_{i=1}^{n_c} \frac{y_i^t}{\pi} g_i}{\sum_{i=1}^{n_c} \frac{y_i^{t-4}}{\pi} g_i} \cong \hat{G}_{olp} = \frac{\sum_{i=1}^{n_c} y_i^t}{\sum_{i=1}^{n_c} y_i^{t-4}}.$$

where there is equality if the correction factor does not exhibit any variability. The variability of the correction factor depends on the variability of the calibration variable (X). We remember that the corrective factor g_i of the initial weight for the i-th unit, is given by:

$$g_i = 1 + (X - \hat{X}_{HT}) \left(\sum_{i=1}^{n} \frac{x_i^2}{\pi_h c_i} \right)^{-1} \frac{x_i}{c_i},$$

where x_i is the value of the calibration variable associated with the i-th unit, X is the true value of its total and $\hat{X}_{HT}$ is its Horvitz-Thompson estimator. As we can see from the standard deviation values in the tables, the improvement of $\hat{G}_{olp.cal}$ compared to $\hat{G}_{olp}$ is higher when the variability of Y^t(and consequently of X and g) increases. In fact, in the first simulation, where the variability of Y^t is quite small, the standard deviation values of $\hat{G}_{olp.cal}$ and $\hat{G}_{olp}$ are very often the same (see

Table 3.7 and 3.10). Instead, in the third simulation, where the variability of Y^t is higher, the standard deviation values of $\hat{G}_{olp.cal}$ are smaller than the $\hat{G}_{olp}$ estimator (see Table 3.9 and 3.12).Obviously, in the case of absence of correlation between the variable of interest and the calibration variable (rho = 0), the results on the standard deviations of the estimators are the same, whether using the calibration or not.

When using calibration, in addition to a smaller standard deviation, an higher overlap value is needed to obtain better results with the estimator that uses only overlapping data. As we can see from Tables 3.7-3.12, this overlap value increases when the rho value increases: to a higher rho value corresponds a higher overlap value over which $Se(\hat{G}_{olp.cal}) < Se(\hat{G}_{all.cal})$. This threshold also increases when the correlation between $Y^t$and Y^{t-4} decreases. In fact, if we compare the just mentioned tables, from simulation 1 to 3, we can notice that the colored part becomes gradually larger. The higher threshold is observed in simulation 3 (Tables 3.9 and 3.12).

For the estimators $\hat{G}_{all}$ and $\hat{G}_{olp}$ (without calibration), the results for the standard deviation are in accordance with those listed in Table 2.2 of the previous chapter, which provides the overlap threshold over which the estimator $\hat{d}_{olp}$ is better than the estimator $\hat{d}_{all}$, for different correlation values between $Y^t$and Y^{t-4}. As in Table 2.2, the results of the three simulations show that when the correlation between Y^t and Y^{t-4} decreases, there is an increase of the overlap threshold over which the estimator using only the overlap data is better than the estimator using all data available in both quarters (see Table 3.6).

Table 3.6 – Overlap threshold over which the estimator $\hat{G}_{olp}$ is better than $\hat{G}_{all}$. Results obtained from the 3 simulations

Simulation	$\rho(y^t, y^{t-4})$	Overlap
Simulation 1: *ε ~ N(0, 0.15)*	0.97	> 0.03
Simulation 2: *ε ~ N(0, 0.25)*	0.92	> 0.09
Simulation 3: *ε ~ N(0, 0.35)*	0.86	> 0.17

Table 3.7 – Theoretical Standard deviation for the estimation of the growth rate g. Simulation 1: ε ~ N(0, 0.15), cor(x,y)=0.97

overlap	rho=0				rho=0.5			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	6.6	5.1	6.6	5.2	5.8	5.0	6.6	5.2
0.10	6.4	3.7	6.4	3.6	5.5	3.6	6.4	3.6
0.15	6.3	3.0	6.3	3	5.4	2.9	6.3	3
0.20	6.1	2.6	6.1	2.6	5.3	2.6	6.1	2.6
0.25	5.9	2.3	5.9	2.3	5.1	2.3	5.9	2.3
0.30	5.7	2.1	5.7	2.1	5.0	2.1	5.7	2.1
0.50	4.9	1.6	4.9	1.6	4.2	1.6	4.9	1.6
0.70	3.8	1.4	3.8	1.4	3.3	1.4	3.8	1.4
0.99	1.3	1.1	1.3	1.1	1.3	1.1	1.3	1.1
o	0.03		0.03		0.04		0.03	

overlap	rho=0.6				rho=0.7			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	5.3	5.0	6.6	5.2	4.7	5.0	6.6	5.2
0.10	5.2	3.6	6.4	3.6	4.6	3.6	6.4	3.6
0.15	5.0	2.9	6.3	3	4.6	2.9	6.3	3
0.20	4.9	2.5	6.1	2.6	4.4	2.5	6.1	2.6
0.25	4.8	2.3	5.9	2.3	4.3	2.3	5.9	2.3
0.30	4.6	2.1	5.7	2.1	4.1	2.1	5.7	2.1
0.50	4.0	1.6	4.9	1.6	3.5	1.6	4.9	1.6
0.70	3.1	1.3	3.8	1.4	2.8	1.3	3.8	1.4
0.99	1.2	1.1	1.3	1.1	1.2	1.1	1.3	1.1
o	0.05		0.03		0.06		0.03	

overlap	rho=0.8				rho=0.9			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	4.0	5.0	6.6	5.2	3.0	5.0	6.6	5.2
0.10	4.0	3.6	6.4	3.6	3.0	3.6	6.4	3.6
0.15	3.9	2.9	6.3	3	2.9	2.9	6.3	3
0.20	3.7	2.5	6.1	2.6	2.8	2.5	6.1	2.6
0.25	3.6	2.3	5.9	2.3	2.7	2.2	5.9	2.3
0.30	3.5	2.0	5.7	2.1	2.6	2.0	5.7	2.1
0.50	3.0	1.6	4.9	1.6	2.3	1.6	4.9	1.6
0.70	2.4	1.3	3.8	1.4	1.9	1.3	3.8	1.4
0.99	1.2	1.1	1.3	1.1	1.1	1.1	1.3	1.1
o	0.08		0.03		0.15		0.03	

overlap	rho=0.95				rho=1			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	2.3	4.9	6.6	5.2	1.1	4.9	6.6	5.2
0.10	2.3	3.6	6.4	3.6	1.1	3.5	6.4	3.6
0.15	2.2	2.9	6.3	3	1.1	2.9	6.3	3
0.20	2.1	2.5	6.1	2.6	1.1	2.5	6.1	2.6
0.25	2.1	2.2	5.9	2.3	1.1	2.2	5.9	2.3
0.30	2.1	2.0	5.7	2.1	1.1	2.0	5.7	2.1
0.50	1.8	1.6	4.9	1.6	1.1	1.6	4.9	1.6
0.70	1.6	1.3	3.8	1.4	1.1	1.3	3.8	1.4
0.99	1.1	1.1	1.3	1.1	1.1	1.1	1.3	1.1
o	0.29		0.03		1.0		0.03	

Table 3.8 – Theoretical Standard deviation for the estimation of the growth rate g. Simulation 2: ε ~ N(0, 0.25), cor(x,y)=0.92

overlap	rho=0				rho=0.5			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	6.7	8.5	6.7	8.7	5.8	8.4	6.7	8.7
0.10	6.6	6.1	6.5	6.1	5.7	6.0	6.5	6.1
0.15	6.4	4.0	6.4	5.0	5.6	4.9	6.4	5.0
0.20	6.2	4.3	6.2	4.3	5.5	4.2	6.2	4.3
0.25	6.0	3.9	6.0	3.9	5.3	3.8	6.0	3.9
0.30	5.9	3.5	5.8	3.5	5.2	3.5	5.8	3.5
0.50	5.1	2.7	5.0	2.7	4.5	2.7	5.0	2.7
0.70	4.1	2.3	4.1	2.3	3.7	2.3	4.1	2.3
0.99	2.0	1.9	2.0	1.9	2.0	1.9	2.0	1.9
o	0.09		0.09		0.11		0.09	
overlap	**rho=0.6**				**rho=0.7**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	5.5	8.4	6.7	8.7	4.9	8.2	6.7	8.7
0.10	5.3	6.0	6.5	6.1	4.9	5.9	6.5	6.1
0.15	5.2	4.9	6.4	5.0	4.7	4.8	6.4	5.0
0.20	5.1	4.2	6.2	4.3	4.6	4.2	6.2	4.3
0.25	4.9	3.8	6.0	3.9	4.5	3.7	6.0	3.9
0.30	4.8	3.4	5.8	3.5	4.4	3.4	5.8	3.5
0.50	4.2	2.7	5.0	2.7	3.8	2.6	5.0	2.7
0.70	3.5	2.2	4.1	2.3	3.2	2.2	4.1	2.3
0.99	1.9	1.9	2.0	1.9	1.9	1.9	2.0	1.9
o	0.13		0.09		0.16		0.09	
overlap	**rho=0.8**				**rho=0.9**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	4.3	8.1	6.7	8.7	3.3	8.1	6.7	8.7
0.10	4.2	5.9	6.5	6.1	3.3	5.8	6.5	6.1
0.15	4.1	4.7	6.4	5.0	3.2	4.7	6.4	5.0
0.20	4.0	4.1	6.2	4.3	3.1	4.1	6.2	4.3
0.25	3.9	3.7	6.0	3.9	3.1	3.6	6.0	3.9
0.30	3.8	3.4	5.8	3.5	3.0	3.3	5.8	3.5
0.50	3.4	2.6	5.0	2.7	2.7	2.6	5.0	2.7
0.70	2.8	2.2	4.1	2.3	2.4	2.2	4.1	2.3
0.99	1.9	1.8	2.0	1.9	1.8	1.8	2.0	1.9
o	0.22		0.09		0.4		0.09	
overlap	**rho=0.95**				**rho=1**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	2.7	8.0	6.7	8.7	1.8	7.9	6.7	8.7
0.10	2.7	5.7	6.5	6.1	1.8	5.7	6.5	6.1
0.15	2.6	4.6	6.4	5.0	1.8	4.6	6.4	5.0
0.20	2.6	4.1	6.2	4.3	1.8	4.0	6.2	4.3
0.25	2.5	3.6	6.0	3.9	1.8	3.6	6.0	3.9
0.30	2.5	3.3	5.8	3.5	1.8	3.3	5.8	3.5
0.50	2.3	2.5	5.0	2.7	1.8	2.5	5.0	2.7
0.70	2.1	2.2	4.1	2.3	1.8	2.1	4.1	2.3
0.99	1.8	1.8	2.0	1.9	1.8	1.8	2.0	1.9
o	0.79		0.09		1		0.09	

Table 3.9 – Theoretical Standard deviation for the estimation of the growth rate g. Simulation 3: ε ~ N(0, 0.35), cor(x,y)=0.86

overlap	rho=0				rho=0.5			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	7	12.1	6.9	12.3	6.1	11.8	6.9	12.3
0.10	6.8	8.7	6.8	8.7	6	8.4	6.8	8.7
0.15	6.6	7.1	6.6	7.1	5.9	6.9	6.6	7.1
0.20	6.5	6.2	6.4	6.1	5.8	6	6.4	6.1
0.25	6.3	5.5	6.2	5.4	5.6	5.3	6.2	5.4
0.30	6.1	5	6.1	5	5.4	4.8	6.1	5
0.50	5.4	3.9	5.3	3.8	4.8	3.8	5.3	3.8
0.70	4.5	3.3	4.4	3.2	4.1	3.1	4.4	3.2
0.99	2.8	2.7	2.8	2.7	2.7	2.6	2.8	2.7
o	0.17		0.17		0.22		0.17	
overlap	**rho=0.6**				**rho=0.7**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	5.7	11.6	6.9	12.3	5.3	11.4	6.9	12.3
0.10	5.6	8.3	6.8	8.7	5.1	8.2	6.8	8.7
0.15	5.5	6.7	6.6	7.1	5.1	6.7	6.6	7.1
0.20	5.4	5.9	6.4	6.1	4.9	5.8	6.4	6.1
0.25	5.2	5.2	6.2	5.4	4.7	5.2	6.2	5.4
0.30	5.1	4.8	6.1	5	4.7	4.7	6.1	5
0.50	4.6	3.7	5.3	3.8	4.2	3.6	5.3	3.8
0.70	3.9	3.1	4.4	3.2	3.7	3.1	4.4	3.2
0.99	2.6	2.6	2.8	2.7	2.6	2.6	2.8	2.7
o	0.25		0.17		0.3		0.17	
overlap	**rho=0.8**				**rho=0.9**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	4.6	11.1	6.9	12.3	3.7	11	6.9	12.3
0.10	4.5	8	6.8	8.7	3.7	7.8	6.8	8.7
0.15	4.4	6.5	6.6	7.1	3.6	6.4	6.6	7.1
0.20	4.3	5.7	6.4	6.1	3.6	5.5	6.4	6.1
0.25	4.2	5	6.2	5.4	3.5	4.9	6.2	5.4
0.30	4.1	4.6	6.1	5	3.4	4.5	6.1	5
0.50	3.7	3.6	5.3	3.8	3.2	3.5	5.3	3.8
0.70	3.3	3	4.4	3.2	2.9	3	4.4	3.2
0.99	2.5	2.5	2.8	2.7	2.5	2.4	2.8	2.7
o	0.41		0.17		0.74		0.17	
overlap	**rho=0.95**				**rho=1**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	3.1	10.8	6.9	12.3	2.4	10.6	6.9	12.3
0.10	3.1	7.8	6.8	8.7	2.4	7.6	6.8	8.7
0.15	3.1	6.3	6.6	7.1	2.4	6.2	6.6	7.1
0.20	3	5.5	6.4	6.1	2.4	5.4	6.4	6.1
0.25	3	4.9	6.2	5.4	2.4	4.8	6.2	5.4
0.30	2.9	4.4	6.1	5	2.4	4.4	6.1	5
0.50	2.8	3.4	5.3	3.8	2.4	3.4	5.3	3.8
0.70	2.7	2.9	4.4	3.2	2.4	2.9	4.4	3.2
0.99	2.4	2.4	2.8	2.7	2.4	2.4	2.8	2.7
o	1		0.17		1		0.17	

Table 3.10 – Standard deviation calculated on 1000 sample estimates for the growth rate g. Simulation 1: ε ~ N(0, 0.15), cor(x,y)=0.97

overlap	rho=0				rho=0.5			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	6.7	4.8	6.6	4.8	6.0	4.7	6.8	4.7
0.10	6.5	3.5	6.3	3.5	5.9	3.6	6.6	3.6
0.15	6.4	2.9	6.2	2.9	5.5	2.8	6.2	2.8
0.20	6.3	2.5	6.1	2.5	5.5	2.6	6.1	2.6
0.25	5.9	2.4	5.7	2.4	5.0	2.2	5.8	2.2
0.30	6.0	2.2	5.8	2.2	5.0	2.1	5.6	2.1
0.50	5.0	1.6	4.9	1.6	4.3	1.6	4.8	1.6
0.70	3.9	1.4	3.8	1.4	3.5	1.4	3.9	1.4
0.99	1.3	1.1	1.3	1.1	1.3	1.2	1.3	1.2
o	0.03		0.08		0.04		0.08	
overlap	rho=0.6				rho=0.7			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	5.5	4.9	6.5	4.9	5.3	4.8	6.7	4.8
0.10	5.4	3.6	6.6	3.6	4.9	3.4	6.1	3.5
0.15	5.3	2.9	6.4	2.9	4.8	3.0	6.2	3.0
0.20	5.1	2.5	6.3	2.5	4.7	2.7	6.1	2.7
0.25	4.7	2.3	5.6	2.3	4.6	2.2	5.8	2.2
0.30	4.6	2.1	5.5	2.1	4.4	2.1	5.6	2.1
0.50	4.0	1.5	4.8	1.5	3.9	1.6	5.0	1.6
0.70	3.3	1.3	4.0	1.3	3.1	1.4	3.8	1.4
0.99	1.3	1.2	1.3	1.2	1.3	1.1	1.4	1.1
o	0.05		0.08		0.06		0.08	
overlap	rho=0.8				rho=0.9			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	4.8	4.9	6.8	4.9	3.8	4.9	6.8	4.9
0.10	4.6	3.5	6.5	3.6	3.5	3.6	6.6	3.6
0.15	4.4	2.9	6.3	2.9	3.3	3.0	6.1	3.0
0.20	4.3	2.5	6.1	2.5	3.3	2.5	5.9	2.6
0.25	4.1	2.2	5.8	2.2	3.2	2.2	5.8	2.2
0.30	4.0	2.0	5.8	2.0	3.3	2.1	5.7	2.1
0.50	3.4	1.5	4.7	1.6	2.7	1.6	4.7	1.6
0.70	2.7	1.3	3.8	1.3	2.3	1.4	3.8	1.4
0.99	1.2	1.1	1.3	1.1	1.1	1.1	1.3	1.1
o	0.08		0.08		0.15		0.08	
overlap	rho=0.95				rho=1			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	2.9	4.9	6.7	5.0	1.1	4.7	6.6	4.8
0.10	2.9	3.4	6.4	3.5	1.1	3.4	6.3	3.5
0.15	2.6	2.9	5.9	3.0	1.1	3.0	6.0	3.0
0.20	2.7	2.5	6.1	2.5	1.1	2.5	6.1	2.5
0.25	2.5	2.3	5.7	2.3	1.1	2.2	5.8	2.2
0.30	2.5	2.0	6.0	2.1	1.1	2.1	5.6	2.1
0.50	2.4	1.6	5.1	1.6	1.1	1.6	4.8	1.7
0.70	1.8	1.4	3.8	1.4	1.1	1.3	3.9	1.3
0.99	1.1	1.1	1.3	1.1	1.1	1.1	1.4	1.1
o	0.29		0.08		1.0		0.08	

Table 3.11 – Standard deviation calculated on 1000 sample estimates for the growth rate g. Simulation 2: ε ~ N(0, 0.25), cor(x,y)=0.92

overlap	rho=0				rho=0.5			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	7.2	8.2	7.0	8.2	5.9	8.0	6.6	8.0
0.10	6.7	5.9	6.4	5.9	5.9	6.0	6.5	6.0
0.15	6.4	4.7	6.2	4.7	5.8	4.8	6.4	4.8
0.20	6.3	4.3	6.1	4.3	5.5	4.2	6.1	4.2
0.25	6.1	3.7	5.9	3.7	5.2	3.8	5.8	3.8
0.30	6.0	3.4	5.8	3.4	5.1	3.5	5.6	3.5
0.50	5.2	2.7	5.1	2.7	4.6	2.6	5.2	2.6
0.70	4.3	2.3	4.1	2.3	3.8	2.2	4.2	2.2
0.99	2.0	1.9	2.0	1.9	2.0	1.9	2.0	1.9
o	0.09		0.09		0.11		0.09	
overlap	rho=0.6				rho=0.7			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	5.6	8.1	6.7	8.1	5.1	8.2	6.4	8.2
0.10	5.6	5.8	6.6	5.8	5.3	5.8	6.6	5.9
0.15	5.5	4.9	6.5	4.9	5.2	4.6	6.5	4.7
0.20	4.9	4.2	5.8	4.2	4.9	4.2	6.2	4.2
0.25	5.2	3.8	6.1	3.9	4.8	3.9	6.0	3.9
0.30	5.1	3.4	5.9	3.5	4.8	3.6	6.0	3.6
0.50	4.4	2.7	5.2	2.7	4.0	2.7	4.9	2.7
0.70	3.7	2.2	4.2	2.2	3.3	2.3	4.1	2.3
0.99	1.9	1.9	2.0	1.9	2.0	2.0	2.1	2.0
o	0.13		0.09		0.16		0.09	
overlap	rho=0.8				rho=0.9			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	4.7	8.2	6.5	8.4	4.0	7.8	6.8	8.0
0.10	4.7	6.1	6.7	6.1	4.0	5.7	6.6	5.8
0.15	4.5	4.9	6.3	4.9	3.9	4.5	6.4	4.6
0.20	4.4	4.1	6.2	4.1	3.8	4.3	6.2	4.4
0.25	4.6	3.8	6.4	3.9	3.6	3.7	6.0	3.7
0.30	4.0	3.4	5.6	3.4	3.4	3.4	5.8	3.4
0.50	3.6	2.7	4.9	2.7	3.2	2.7	5.2	2.8
0.70	3.1	2.2	4.1	2.2	2.8	2.2	4.2	2.3
0.99	1.9	1.9	2.0	1.9	1.9	1.9	2.0	1.9
o	0.22		0.09		0.40		0.09	
overlap	rho=0.95				rho=1			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	3.0	7.9	6.5	8.1	1.8	7.8	7.0	8.1
0.10	3.0	6.0	6.6	6.2	1.8	5.8	6.4	6.0
0.15	3.1	4.7	6.6	4.9	1.8	4.6	6.7	4.8
0.20	3.0	4.2	6.1	4.3	1.8	4.0	6.2	4.3
0.25	2.9	3.8	6.2	3.9	1.8	3.7	6.1	3.9
0.30	2.8	3.4	5.9	3.5	1.8	3.4	5.7	3.6
0.50	2.5	2.6	5.0	2.7	1.8	2.6	5.0	2.7
0.70	2.3	2.3	4.0	2.3	1.8	2.2	4.0	2.3
0.99	1.7	1.7	1.9	1.8	1.7	1.8	2.0	1.9
o	0.79		0.09		1.00		0.09	

Table 3.12– Standard deviation calculated on 1000 sample estimates for the growth rate g. Simulation 3: ε ~ N(0, 0.35), cor(x,y)=0.86

overlap	rho=0				rho=0.5			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	7.2	11.5	6.9	12.3	6.1	11.4	6.9	12.3
0.10	7.1	8.6	6.8	8.7	6.1	8.2	6.8	8.7
0.15	6.8	7.1	6.6	7.1	5.9	6.9	6.6	7.1
0.20	6.7	6.3	6.4	6.1	6	6.1	6.4	6.1
0.25	6.3	5.5	6.2	5.4	5.8	5.4	6.2	5.4
0.30	6.3	4.8	6.1	5	5.6	4.8	6.1	5
0.50	5.4	3.8	5.3	3.8	4.9	3.8	5.3	3.8
0.70	4.5	3.2	4.4	3.2	4.2	3.2	4.4	3.2
0.99	2.8	2.7	2.8	2.7	2.7	2.7	2.8	2.7
o	0.17		0.17		0.22		0.17	
overlap	rho=0.6				rho=0.7			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	5.9	11.5	6.9	12.3	5.8	11.3	6.9	12.3
0.10	5.8	8.4	6.8	8.7	5.8	8.4	6.8	8.7
0.15	5.8	6.8	6.6	7.1	5.4	6.6	6.6	7.1
0.20	5.7	6.1	6.4	6.1	5.2	5.7	6.4	6.1
0.25	5.4	5.3	6.2	5.4	5	5.4	6.2	5.4
0.30	5.5	5	6.1	5	5.1	4.8	6.1	5
0.50	4.7	3.8	5.3	3.8	4.3	3.9	5.3	3.8
0.70	3.8	3	4.4	3.2	3.8	3.2	4.4	3.2
0.99	2.6	2.6	2.8	2.7	2.8	2.7	2.8	2.7
o	0.25		0.17		0.30		0.17	
overlap	rho=0.8				rho=0.9			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	5.2	11.8	6.9	12.3	4.5	11.3	6.9	12.3
0.10	4.8	8	6.8	8.7	4.4	8.2	6.8	8.7
0.15	4.8	6.9	6.6	7.1	4.1	6.7	6.6	7.1
0.20	4.9	5.9	6.4	6.1	4.2	5.8	6.4	6.1
0.25	4.8	5.1	6.2	5.4	4	5.4	6.2	5.4
0.30	4.6	4.9	6.1	5	3.8	4.7	6.1	5
0.50	4.3	3.8	5.3	3.8	3.6	3.8	5.3	3.8
0.70	3.6	3.3	4.4	3.2	3.2	3.2	4.4	3.2
0.99	2.5	2.5	2.8	2.7	2.4	2.4	2.8	2.7
o	0.41		0.17		0.74		0.17	
overlap	rho=0.95				rho=1			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	3.7	11.1	6.9	12.3	2.4	10.8	6.9	12.3
0.10	3.5	7.9	6.8	8.7	2.4	8	6.8	8.7
0.15	3.6	6.7	6.6	7.1	2.4	6.1	6.6	7.1
0.20	3.5	5.8	6.4	6.1	2.4	5.4	6.4	6.1
0.25	3.3	5.1	6.2	5.4	2.5	4.7	6.2	5.4
0.30	3.6	4.8	6.1	5	2.5	4.7	6.1	5
0.50	3.3	3.6	5.3	3.8	2.5	3.4	5.3	3.8
0.70	3	3.1	4.4	3.2	2.5	2.9	4.4	3.2
0.99	2.5	2.5	2.8	2.7	2.4	2.4	2.8	2.7
o	1.0		0.17		1.0		0.17	

For each overlap and rho value, empirical bias, mean squared error have been computed, and confidence intervals for the estimates of g obtained.

The empirical bias is computed as the difference between the mean of the growth rate obtained from the 1000 estimates of g and his true value. The results are showed in the Appendix (Tables 1-3). The absolute bias calculated from the 1000 estimates is very small. In fact, for most cases the bias is approximately equal to 0.

For each estimate, a t-Student distribution was used, and the corresponding 95% confidence intervals were calculated, according the following formulas

$$I(\hat{g}_{all.cal}) = \hat{g}_{all.cal} \pm t_{n_r-1}^{0.975} Se(\hat{g}_{all.cal})$$

$$I(\hat{g}_{olp.cal}) = \hat{g}_{olp.cal} \pm t_{n_c-1}^{0.975} Se(\hat{g}_{olp.cal})$$

$$I(\hat{g}_{all}) = \hat{g}_{all} \pm t_{n_r-1}^{0.975} Se(\hat{g}_{all})$$

$$I(\hat{g}_{olp}) = \hat{g}_{olp} \pm t_{n_c-1}^{0.975} Se(\hat{g}_{olp}).$$

The actual coverage probability of such confidence intervals is computed via simulation as the proportion of simulated confidence intervals that contain the true value of the growth rate g. The results are shown in Appendix (Tables 7-9). As expected, the actual coverage probability is close to its nominal value, i.e. 95%. However, smaller values are obtained if the $\hat{g}_{olp}$ and the $\hat{g}_{olp.cal}$ estimators are used. In this case, especially for small overlap levels (5-10%), the coverage probability is approximately 90%. This is due to the fact that with low levels of overlap, the estimates were calculated on a small number of units (n_c). For example, with an overlap of 5%, only 15 units were used for the estimation.

Finally, in Tables 10-15 in the Appendix are shown the coefficients of variation for the estimates of the totals Y^t and Y^{t-4} obtained using calibration. As expected, for each simulation, the coefficients of variation for the totals are always smaller using the $\hat{g}_{all.cal}$ estimator rather than the $\hat{g}_{olp.cal}$ estimator.

3.2 - Simulation in the case of stratified population

3.2.1 - Aim and main steps of the simulation study

This paragraph focuses on the estimate of the change in case of a stratified population. The main step followed in this simulation exercise are:

- A population of $N = 19{,}889$ units has been generated with the turnover having a lognormal distribution with parameters (mean and variance) that reproduce the population observed within each stratum in the sector of Accommodation. The population is divided into four strata based on the size of the company. The population in the previous simulation and the population within the stratum 1 in this simulation are generated according the same distribution and parameters.

- The sample size is calculated by means of the Bethel algorithms implemented in Mauss-R (see Barcaroli et al, 2010). The planned coefficient of variation for the estimation of the total turnover is fixed at 3% for the estimation domain. The sample size (n) obtained within the estimation domain is 388 units, with a sampling fraction of 2%.

- For the generation of the population at the next occasion (t) it has been assumed the following behavior:

$$Y_i^t = \beta Y_i^{t-4} + \varepsilon_i \, Y_i^{t-4}$$
$$\varepsilon \sim N(0, 0.15)$$

  As in previous paragraph, the value of β has been fixed at 0.9 and ε_i is a random variable having a normal distribution.

- The sample at the occasion t-4 is extracted from the reference population. A random non-response of 30% of the units in the sample is applied. As a consequence, the size of the set of respondent units for the estimation domain is equal to $n_r = 272$.

- The sample at the occasion t consists in the union of a random subset of the respondent units at the occasion t-4, with size n_c depending on the desired overlapping o $(n_c = on_r)$ and of a srswor of size $n_r - n_c$ from the population (units in the first subset are excluded). Therefore the size of respondent units is the same for both occasions. In this simulation the value of the overlap o is fixed to 0.7.

- The estimates of the growth rate of the total turnover in the population between the two occasions are estimated by using the four estimators described.
- The estimates are computed on 300 different samples extracted from the reference population. This allows the calculation of the bias, the standard deviation and the mean squared error for each estimator used.

Table 3.13 contains the summary statistics about the generated population for the occasion *t* and *t-4*. There is a strong correlation between Y^t and Y^{t-4} (0.98 within the estimation domain).

Table 3.13 – Summary statistics of the simulation exercise in case of stratification of the population

Strata	N	n	Sampling fraction%	n_r	o	n_c	$cor(Y^t, Y^{t-4})$	percentage growth rate g
1	8,413	30	0.4	21	0.7	14	0.98	-10.1
2	9,885	140	1.4	98	0.7	69	0.97	-9.8
3	1,456	83	5.7	58	0.7	41	0.98	-9.6
4	135	135	100	95	0.7	66	0.95	-9.2
Tot.	19,889	388	2.0	272	0.7	190	0.98	-9.7

3.2.2 - Results of the simulation

The bias, the standard deviation and the mean squared error within strata were calculated using the same methodology of the simulation described in Section 3.1.

The results obtained within the stratum 1 in the current simulation may be compared with those obtained in the previous simulation (where *ε ~ N(0, 0.15)*, overlap=0.7 and rho=0.95) because the populations are generated according to the same distribution and parameters. It is seen that both bias and standard deviation in stratum 1 are larger than those obtained in the previous simulation. This is because in the present case the sample error was set at 3% on the entire estimation domain. Consequently, the sample size in stratum 1 is considerably smaller than the one obtained in the previous simulation (30 vs 417).

The $\hat{G}_{d,olp}$ and $\hat{G}_{d,all}$ estimators were calculated using the methodology described in the Section 2.2.1. Therefore, we have

$$\hat{I}_d = \sum_{h=1}^{H} \hat{I}_h w_h$$

$$\hat{G}_{d,olp} = \frac{\hat{I}_d^t}{I_d^{t-4}}.$$

The strata indices referring to the first occasion have been set equal to 100, while the strata weights were computed from the Istat Statistic Register of active firms (ASIA), used also to compute the lognormal distribution parameters for the generation of the population.

As we can see from Tables 3.14, 3.15 the estimators have a strong bias and standard deviation within the strata. Stratum 4 is an exception, because it is a census stratum. Instead, within the estimation domain the bias is nearly 0 for all the estimators except for the estimator G_{all} (1.1 p.p.). Standard deviations within the estimation domain are smaller than the ones within the strata.

The best estimators are $\hat{G}_{olp.cal}$ and $\hat{G}_{olp}$. For these estimators, the mean squared error within the estimation domain is the same. This is probably due to the low variability of the calibration variable within the strata, which makes the calibrated weights very similar each other. Therefore we have that $\hat{G}_{olp.cal} \cong \hat{G}_{olp}$.

Table 3.14 – Bias (p.p) calculated on 300 sample estimates for the growth rate g.

Simulation: ε ~ N(0, 0.15), o=0.7, rho=0.95

Stratum/ Domain	calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp
Stratum 1	1.5	-0.4	4.7	-0.5
Stratum 2	0.1	-0.2	0.7	-0.2
Stratum 3	0.4	0.1	1.6	0.2
Stratum 4	0	0	0	0
Domain	0.2	-0.1	1.1	-0.2

Table 3.15 – Standard deviation calculated on 300 sample estimates for the growth rate g. Simulation: ε ~ N(0, 0.15), o=0.7, rho=0.95

Stratum/ Domain	calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp
Stratum 1	14.7	5.7	26.2	5.9
Stratum 2	4.2	2.7	8.7	2.8
Stratum 3	5.4	3.1	11.2	3.2
Stratum 4	2.2	1.8	5.2	1.8
Domain	2.8	1.5	5.4	1.5

Table 3.16 – Mean squared error calculated on 300 sample estimates for the growth rate g. Simulation: ε ~ N(0, 0.15), o=0.7, rho=0.95

Stratum/ Domain	calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp
Stratum 1	219.5	32.7	708.9	35.0
Stratum 2	17.9	7.4	76.2	7.9
Stratum 3	29.5	9.8	128.0	10.3
Stratum 4	4.7	3.3	27.0	3.2
Domain	8.1	2.3	30.4	2.3

From the present simulation study we can deduce that for the sector of Accommodation, in the case of:

- a very high correlation between Y^t and Y^{t-4},
- an overlap between the two occasions near to 0.7,
- a correlation between Y^t and the calibration variable near to 0.95,

the best estimators for the growth rate are the estimators $\hat{G}_{olp.cal}$ and $\hat{G}_{olp}$. As a consequence, for the growth rate estimation it is better to use only the overlapping respondent units than all the respondent units of the two reference quarters.

In the next chapter we will see an application to real data for different estimation domains.

CHAPTER 4

Application to real data

4.1 - An application to the service turnover survey data

This chapter describes an application performed on real data using information from the quarterly service turnover survey. The application was performed on 2 different domains of the Nace Rev.2 Classification (two-digit numerical code).

The first domain (D1) consists of four different estimation domains (G1, G2, G3, G4) that match with the groups of the Nace Rev.2 Classification (three-digit numerical code). The second domain (D2) consists of two different estimation domains (G5 and G6), where G5 is a group of the Nace Rev.2 classification while G6 is an aggregation of the other groups within the domain D2.

Each domain estimation (G1, G2, G3, G4, G5 and G6) is divided into four independent strata according to the class of employees, with the exception of one estimation domain (G1), which is instead divided into three independent strata. The stratum with more than 100 employees within each estimation domain is the self-representative stratum. The application has been conducted on a given estimation quarter (which is not specified here).

The estimators used for the growth rate estimation are those described in the previous chapter ($\hat{g}_{d,olp}, \hat{g}_{d,all.cal}, \hat{g}_{d,olp.cal}$). Since, as seen from the simulation study, the estimator $\hat{g}_{d,all}$ gives the worst results in terms of standard error of the growth rate estimation, it has not been used in the present application. Therefore, in the application are used the same estimators of the service turnover survey ($\hat{g}_{d,olp}, \hat{g}_{d,all.cal}$) and the estimator $\hat{g}_{d,olp.cal}$ (non currently used in the service turnover survey).

The results for the growth rate estimations are showed in Table 4.1. As already mentioned in Chapter 2, the sample size n is calculated by means of the Bethel algorithm using the software Mauss-R and with the planned coefficient of variation fixed at 3% for the total estimation within each estimation domain. In Table 4.1 the following quantities are shown.

- Weight (w) of each group within the domains of the Nace Rev.2 Classification, in terms of turnover. It is calculated from the Istat Statistic Register of Active Firms (ASIA) and it is necessary for the calculation of the growth rate and standard error estimation when the estimator $\hat{g}_{d,olp}$ is used (as showed in the chapter 2).

- Number of units in the population (N), theoretical sample size calculated using the software Mauss-R (n), and corresponding percentage sampling fraction.

- Average response rate of the units in the quarters t and t-4.

Table 4.1 - Growth rate estimation for some estimation domains of the turnover service survey using different estimators.

Group/ Domain	w	N	n	Sampling fraction%	Resp. rate%	$\hat{g}_{d,olp}$	$\hat{g}_{d,all.cal}$	$\hat{g}_{d,olp.cal}$
G1	0.68	3,538	270	7.6	80%	14.2	12.7	13.3
G2	0.13	39,817	522	1.3	77%	-2.0	-1.8	-2.1
G3	0.16	8,763	532	6.0	75%	2.6	2.7	2.5
G4	0.03	2,835	381	13.4	71%	2.3	3.8	2.6
D1	**1**	**54,953**	**1,705**	**3.1**	**76%**	**9.9**	**8.9**	**9.1**
G5	0.83	19,887	475	2.4	67%	3.2	3.7	3.3
G6	0.17	8,135	444	5.5	69%	7.1	7.6	7.2
D2	**1**	**28,022**	**919**	**3.3**	**68%**	**3.9**	**4.5**	**4.2**

As we may see from Table 4.1, the sampling fraction at domain level (D1 and D2) is just over 3%. The percentage response rate is around 70%, (this motivated the choice to apply a 30% of non-response in the simulation study of the previous chapter).

The growth rate estimates performed with the three different estimators vary between 8.9% and 9.9 for the domain D1 and between 3.9% and 4.5% for the domain D2. As mentioned in the chapter 2, when the estimator $\hat{g}_{d,all.cal}$ and $\hat{g}_{d,olp.cal}$ are used, the calibration variable used for the calculation of the totals $\hat{Y}^{t}$ and $\hat{Y}^{t-4}$ is the annual turnover deriving from the last Asia available. The sample correlation between the variable of interest and the calibration variable (*rho*) is very high (0.99 for the domain D1 and 0.96 for the domain D2).

In the next section we evaluate the standard errors associated with the different estimates.

4.2 - Standard error using the Taylor series approximation and a comparison with the bootstrap method

Standard errors have been calculated using the Taylor series approximation. When the calibration was used ($\hat{g}_{d,all.cal}, \hat{g}_{d,olp.cal}$), the results for the standard errors obtained through the Taylor series approximation were compared with those obtained using the bootstrap method (see Efron B. 1982; Rao and Wu 1988; Holmberg A. 1998; Antal and Tillé 2012; Quatember A. 2014).

Using the method proposed by Holmberg (1998), three artificial stratified populations (U_t^*, $U_{t,t-4}^*$ and U_{t-4}^*) were created, by replicating for a certain number of times ($d_{h,k}$,) the value collected on each respondent unit (k) in the quarters t and t- 4 (y_t and y_{t-4}). The artificial populations U_t^* and U_{t-4}^* were created by replicating the values on the units responding only to one of the two quarters (t and t-4 respectively), while the population $U_{t,t-4}^*$ was created by replicating the values on the overlapping respondent units to both quarters.

The number of times that the value of a unit needs to be replicated is given by:

$$\mathrm{d}_{\mathrm{h,k}} = c_{h,k} + \varepsilon_{h,k} \ ,$$

where $c_{h,k}$ is the integer part of the inverse of the probability of inclusion for the units k belonging to the stratum h ($\pi_{h,k}$) and $\varepsilon_{h,k}$ is the realization of a random variable with Bernoulli distribution. The Bernoulli distribution parameter is given by the difference between the inverse of the inclusion probability and its integer part ($r_{h,k}$):

$$c_{h,k} = \lfloor \pi_{h,k}^{-1} \rfloor = \left\lfloor \frac{N_h}{n_h} \right\rfloor$$

$$r_{h,k} = \pi_{h,k}^{-1} - c_{h,k}$$

$$\varepsilon_{h,k} = Ber\left(r_{h,k}\right),$$

Since in these estimation domains a stratified *srswor* is used, the inclusion probabilities ($\pi_{h,k}$) are the same for each unit belonging to the same stratum h.

300 bootstrap samples were generated from the artificial resampling populations in such a way that the overlapping of the units between the two quarters is the same as the parent sample, within each stratum. For each stratum (h) we have:

- a number of units equal to the number of respondent units only in the quarter *t-4* (the units in s^{t-4}) has been extracted from the population U_{t-4}^{*}. These extracted bootstrap units are represented by s_b^{t-4} in the figure 4.1.
- a number of units equal to the number of respondent units only in the quarter t (the units in s^{t}) has been extracted from the population U_t^{*}. These extracted bootstrap units are represented by s_b^{t} in the figure 4.1.
- a number of units equal to the number of respondent units in both quarters (the units in $s^{t,t-4}$) has been extracted from the population $U_{t,t-4}^{*}$. These extracted bootstrap units are represented by $s_b^{t,t-4}$ in the figure 4.1.

The bootstrap sample units in the quarter *t-4* were constituted by the union of $s_b^{t-4}$and $s_b^{t,t-4}$ while the bootstrap sample units in the quarter t were constituted by the union of $s_b^{t,t-4}$ and s_b^{t} (see figure 4.1).

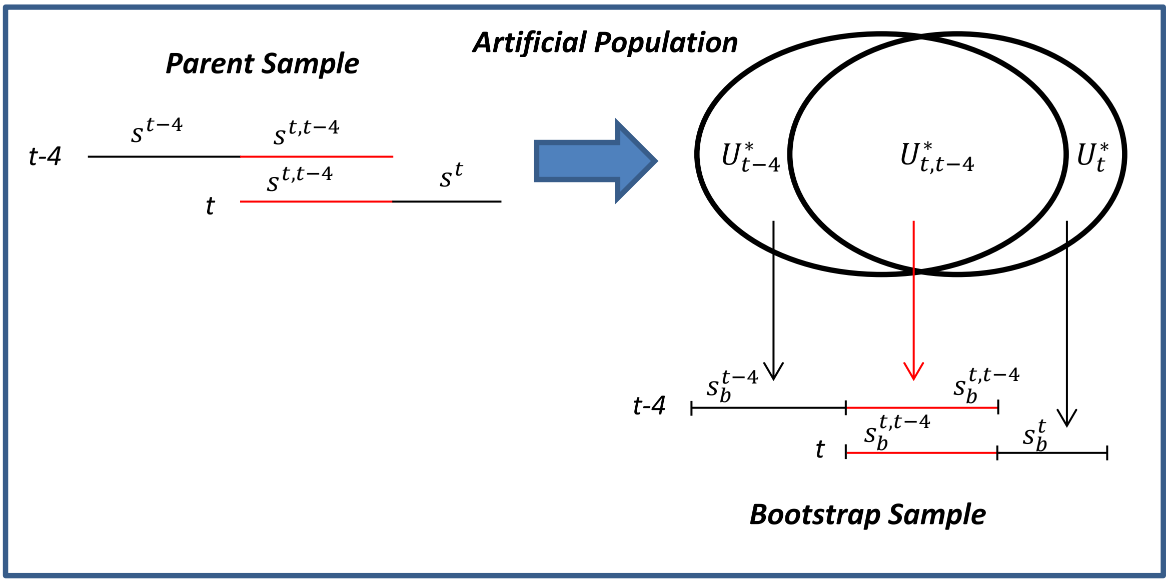
For each of the 300 bootstrap samples, an estimate of the growth rate of the turnover was computed, using the estimators $\hat{g}_{all.cal}$ and $\hat{g}_{olp.cal}$. Afterwards, for the estimation of the standard error was used the Monte Carlo bootstrap variance estimator, obtained by the following formula:

$$\hat{V}_{Bwo} = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{g}_b^* - \bar{\hat{g}}^*\right)^2$$

where B = 300 and:

$$\bar{\hat{g}}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{g}_B^*$$

Figure 4.1 – Creation of the bootstrap sample from the parent sample



The results for the standard error are showed in Table 4.2. Observing the results obtained through the Taylor series approximation, the best results are obtained with the use of the estimator $\hat{g}_{olp.cal}$. Therefore, the use of the calibration only on the respondent units to both quarters gives the best results in terms of the standard error:

- The reason for which $Se(\hat{g}_{olp.cal}) < Se(\hat{g}_{all.cal})$ is that the overlapping rate of the respondents between the two quarters is very high (over 70%), as well as the sample correlation between the variable of interest of the units in the same stratum in the two different occasions ($cor(y^t, y^{t-4}) \cong 0.99$). In fact, as seen in the previous chapter, in our simulation study we have already remarked that at higher correlation levels between Y^t and Y^{t-4} there is a lower overlapping value (o) over which $Se(\hat{g}_{olp.cal}) < Se(\hat{g}_{all.cal})$. In our simulation study, in the case of $cor(y^t, y^{t-4}) = 0.97$ and a correlation between y^t and the calibration variable (*rho*) equal to 0.95, the overlapping value (*o*) over which $Se(\hat{g}_{olp.cal}) < Se(\hat{g}_{all.cal})$ is equal to 29%. These results also show that the difference $Se(\hat{g}_{all.cal}) - Se(\hat{g}_{olp.cal})$ is larger within the D2 domain. This probably happens because the overlapping between the respondents in the two occasions within the D1 domain, is very high. In fact, as we approach the case of full overlapping the results on standard errors tend to converge.

- The reason for which $Se(\hat{g}_{olp.cal}) < Se(\hat{g}_{olp})$ it is probably due to the fact that the calibration improves the precision of the estimates thanks to the high correlation between the variable of interest and the calibration variable. This is true especially within the D1 domain, where the correlation with the calibration variable is higher.

By comparing the standard error of the estimators currently used in the service turnover survey ($\hat{g}_{olp}$ and $\hat{g}_{all.cal}$) we may see that

- $Se(\hat{g}_{all.cal}) < Se(\hat{g}_{olp})$ within the domain D1
- $Se(\hat{g}_{all.cal}) > Se(\hat{g}_{olp})$ within the D2 domain.

This could depend on the fact that the correlation with the calibration variable is greater within the D1 domain than within the D2 domain. Therefore, within the D1 domain, the loss of precision in the use of an estimator based on all respondents rather than an estimator based only on the overlapping respondents is compensated by the use of the calibration with a variable highly correlated to that of interest.

Since in some strata of the domain G5 and G6, the estimation of the covariance term of the Taylor series approximation led to a negative value of $\hat{V}ar(\hat{g}_{olp.cal})$, the covariance term estimation for these domains were made in the following way:

$$\hat{C}ov(\hat{Z}^{t-4}_{h,all.cal}, \hat{Z}^{t}_{h,all.cal}) =$$

$$\hat{c}or\left(\left(z^t_{h,i}q^t_{h,i} - \bar{\tilde{z}}^t_{h,r2}\right)\left(z^{t-4}_{h,i}q^{t-4}_{h,i} - \bar{\tilde{z}}^{t-4}_{h,r2}\right)\right)\sqrt{\hat{V}ar(\hat{Z}^{t-4}_{olp.cal})\hat{V}ar(\hat{Z}^{t}_{olp.cal})},$$

instead of:

$$Cov\left(\hat{Z}^{t-4}_{h,all.cal}, \hat{Z}^{t}_{h,all.cal}\right) = N_h^2\left(\frac{1}{n_{h,r_2}} - \frac{1}{N_h}\right) S_{z_h^t, z_h^{t-4}}$$

The results obtained with the bootstrap method in terms of standard errors are quite close to those obtained with the Taylor series approximation. The main difference is obtained for the estimate of the standard error when the estimator $\hat{g}_{olp}$ is used. In fact, the standard error values for the estimates obtained using the $\hat{g}_{olp}$ estimator, through the bootstrap method, are smaller than those obtained through the Taylor Series Approximation and they are also closer to those obtained with the use of the $\hat{g}_{olp.cal}$ estimator.

Table 4.2 – Standard error of the growth* rate *estimation for some estimation domains of the service turnover survey (three-digit numerical code of the Nace Rev.2 classification).

Domain /Group	Overlap	Taylor series Approximation			Bootstrap method		
		$\hat{S}e$ $(\hat{g}_{olp})$	$\hat{S}e$ $(\hat{g}_{all.cal})$	$\hat{S}e$ $(\hat{g}_{olp.cal})$	$\hat{S}e$ $(\hat{g}_{olp})$	$\hat{S}e$ $(\hat{g}_{all.cal})$	$\hat{S}e$ $(\hat{g}_{olp.cal})$
G1	0.84	1.4	1,3	1,2	1.1	1,1	1,0
G2	0.78	1.4	1,3	1,3	1.3	1,2	1,3
G3	0.82	1.1	1,0	0,7	0.8	0,7	0,7
G4	0.74	1.2	1,3	1,0	1.1	1,2	1,1
D1	**0.79**	**1.0**	**0,9**	**0,8**	**0.8**	**0,8**	**0,7**
G5	0.72	0.9	1,9	0,9	0.8	1,7	0,9
G6	0.70	0.7	1,7	0,7	0.6	1,4	0,7
D2	**0.71**	**0.8**	**1,5**	**0,7**	**07**	**1,4**	**0,7**

However, both methods would seem to suggest that at these levels of sample overlap and correlation with the calibration variable, the best results in terms of standard error are obtained with the use of the $\hat{g}_{olp.cal}$ estimator.

4.3 - A comparison with the Knottnerus and Van Delden results about the standard error of the turnover growth rate in Dutch supermarkets

As mentioned in Chapter 1, Knottnerus and Van Delden (2012) gave the results about the growth rates and their confidence interval at 95% level, of monthly turnover (compared to 12 months previous) in the dutch supermarket, between the years 2003 and 2004. In table 4.3 are showed their results. The confidence intervals are given between parantheses.

Table 4.3 – Estimated growth rates and their 95% margins. Results obtained by Knottnerus and Van Delden for Dutch Supermarkets.

t	$\hat{g}$
16	$-0.3(\pm 1.0)$
17	$-3.7(\pm 1.0)$
18	$1.6(\pm 1.0)$
19	$-2.2(\pm 0.9)$
20	$0.5(\pm 0.8)$
21	$-1.7(\pm 0.8)$
22	$-2.2(\pm 0.8)$
23	$0(\pm 0.8)$
24	$-2.3(\pm 0.9)$

As we can see in the above table, the 95% margins vary between 0.7 and 1.0 per cent point. We compare now these results with those obtained for the D1 and D2 domains in the application described in the previous paragraphs. The results are showed in the table 4.4. The 95% margins, vary between 1.4 and 1.6 per cent point. Compared to the results for monthly turnover in Dutch supermarkets, the standard errors calculated for the turnover within the two domains (D1 and D2) are higher. However, we need to consider the different sampling rates of the two surveys. The sample for the turnover survey in Dutch Supermarkets consists in 900 units out of a population of 3,500 units. Therefore the sampling fraction is of 26%, much higher than the one for the D1 and D2 domains within the Italian turnover (about 3%). This may explain the larger margins obtained for the D1 and D2 domains.

Table 4.3 – Estimated growth rates and their 95% margins within the domain D1 and D2 (two-digit numerical code of the Nace Rev.2 classification)

t	$\hat{g}$
D1	$9.1(\pm 1.6)$
D2	$-4.2(\pm 1.4)$

.

Conclusions

The aim of this work was to compute the variance of the estimators currently used in the service turnover survey for the quarterly estimation of the turnover growth rate.

The survey currently uses two indicators for the estimation of the growth rate. The first one is a ratio between two mean estimators (one for quarter t and one for quarter t-4) and is calculated on the set of respondents common to both quarters (this estimator is indicated with $\hat{G}_{olp}$). The second estimator is instead the ratio between two totals (one for quarter t and one for quarter t-4), calculated using the calibration estimator. This second estimator is applied to the whole set of respondents in both periods, t and t-4 (this estimator is indicated with $\hat{G}_{all.cal}$).

This work had also the purpose to determine which is the best estimator in terms of standard error. Since both estimators are non-linear functions of linear estimators, the first-order Taylor approximation was used to compute the variance. Therefore, it was possible to find the formulation of the variance of these estimators, both at stratum and at domain level.

A simulation study has been conducted: two populations referred to two different occasions (t and t-4) were generated with turnover values at the occasion t-4 possessing a lognormal distribution with parameters (mean and variance) able to reproduce the population observed in the sector of Accommodation. 1,000 samples were extracted from the generated population. Therefore, it was possible to compute the bias, the standard deviation and the mean squared error for the estimation of the turnover growth rate. The analysis was performed for different sample overlapping values between the two reference quarters (t and t-4) and different correlation values between the variable of interest and the calibration variable, together with different correlations between Y^t and Y^{t-4}. Both estimators used in the service turnover survey were applied, as well as two additional estimators: the $\hat{G}_{olp.cal}$ estimator, that was computed on the respondent units common to the two reference quarters using the calibration estimator; and the $\hat{G}_{all}$ estimator which uses the set of all respondent units at the two occasions and is computed on the ratio between the two sample means in two different quarters, like the $\hat{G}_{olp}$ estimator. The simulation study was carried on in case of simple random sampling design and in case of stratified sampling design. The study highlighted that to all estimators, at a higher overlap level between the respondent units for the two occasions, correspond a smaller standard error. For the estimator using calibration ($\hat{G}_{olp.cal}$ e $\hat{G}_{all.cal}$), at a higher value of correlation between the variable of interest and the calibration variable, the standard errors are smaller, while the overlap threshold over which $Se(\hat{G}_{all.cal}) > Se(\hat{G}_{olp.cal})$ is higher.

The study shows that, with the same set of respondents, the results obtained through the use of calibration are better than the ones obtained by using the mean estimators (we have that $Se(\hat{G}_{all.cal}) < Se(\hat{G}_{all})$ and $Se(\hat{G}_{olp.cal}) < Se(\hat{G}_{olp})$). Moreover, the simulation study in case of stratified population shows that, at a level of overlap between the two occasions of about 70%, a correlation between the variable of interest and the calibration variable equal to 0.95 and a very high correlation between the observations in the two different occasions, $\hat{G}_{olp.cal}$ is the estimator with the smaller mean squared error associated to the estimation (the results are very similar to the ones obtained with the estimator $\hat{G}_{olp}$)

In the last part of the work has been conducted an application performed on real data, using information from the quarterly service turnover survey. The confidence intervals associated with the year-over-year variation of the quarterly service turnover were calculated for some estimation domains. The standard errors obtained by using Taylor first-order series approximation were compared with the ones obtained with the bootstrap method. The comparison shows similar results, although it appears that the results obtained with the Taylor series approximation are more conservative, as they are wider. The smallest standard errors were obtained through the use of the $\hat{G}_{olp.cal}$ estimator.

In conclusion, the simulation study and the application show that, given the characteristics of the service turnover survey, the estimator with the smallest standard errors is the calibration estimator calculated only on the overlapping sample units in both quarters ($\hat{G}_{olp.cal}$). The above mentioned characteristics are: a high overlapping level (above 70%), a high correlation between the variable of interest and the calibration variable (greater than 0.95) and a very high correlation between the observations in the two occasions.

Results discussed in the thesis refer to srswor and stratified srswor. Nonetheless, in future research, it may be interesting to extend the approach to more complex sampling designs. Furthermore, possible future developments of the work could be to analyze how the estimators perform for different levels of variation between the two survey occasions; what are the effects on them, on their bias and validity of the corresponding expressions of variance of small sample sizes in the strata and how different non-response mechanisms can influence the choice between them, above all the choice between the use of calibration or not.

Appendix

Table 1 – Bias (p.p) calculated on 1000 sample estimates for the growth rate g. Simulation 1: ε ~ N(0, 0.15), $\boldsymbol{cor}(\boldsymbol{Y^t},\boldsymbol{Y^{t-4}}) = \mathbf{0.97}$

overlap	**rho=0**				**rho=0.5**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0,05	0,5	-0,2	0,5	-0,2	0,6	-0,1	0,6	-0,1
0,10	0,2	0,1	0,2	0,1	0,0	-0,1	0,0	-0,1
0,15	0,4	-0,1	0,3	-0,1	0,0	0,0	0,1	0,0
0,20	-0,2	0,0	-0,2	0,0	0,3	0,1	0,5	0,1
0,25	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,0
0,30	0,0	0,0	0,0	0,0	0,3	0,1	0,4	0,1
0,50	0,1	-0,2	0,1	-0,2	0,2	0,0	0,2	0,0
0,70	-0,1	0,0	-0,1	0,0	0,1	0,0	0,1	0,0
0,99	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
overlap	**rho=0.6**				**rho=0.7**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0,05	0,1	0,0	0,2	0,0	0,2	-0,2	0,3	-0,1
0,10	0,5	0,1	0,6	0,1	0,0	-0,1	0,1	-0,1
0,15	0,2	0,1	0,3	0,1	0,0	0,0	0,1	0,0
0,20	0,1	-0,1	0,0	-0,1	0,0	0,0	0,1	0,0
0,25	0,0	-0,1	0,2	0,0	-0,1	0,0	-0,1	0,0
0,30	0,3	-0,1	0,5	-0,1	0,2	-0,1	0,4	-0,1
0,50	0,1	0,0	0,0	0,0	0,1	0,0	0,2	0,0
0,70	0,1	0,1	0,1	0,1	0,0	-0,1	0,1	-0,1
0,99	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
overlap	**rho=0.8**				**rho=0.9**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0,05	0,2	-0,2	0,2	-0,2	0,0	0,0	0,2	0,1
0,10	0,3	0,2	0,4	0,3	0,0	0,0	0,2	0,0
0,15	0,1	0,0	0,3	0,0	0,1	-0,1	0,2	0,0
0,20	0,0	0,0	0,0	0,0	0,1	-0,1	0,3	-0,1
0,25	-0,1	0,0	-0,1	0,0	0,1	-0,1	0,3	0,0
0,30	0,0	0,0	0,4	0,0	0,0	0,0	0,3	0,0
0,50	-0,1	-0,1	-0,1	-0,1	0,2	0,0	0,4	0,0
0,70	0,0	0,0	0,0	0,0	0,1	0,1	0,1	0,1
0,99	-0,1	-0,1	0,0	-0,1	0,0	-0,1	0,0	-0,1
overlap	**rho=0.95**				**rho=1**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0,05	-0,1	-0,2	0,1	-0,2	-0,1	-0,3	0,2	-0,2
0,10	0,0	0,0	0,2	0,1	0,0	-0,1	-0,1	-0,1
0,15	0,0	0,2	0,1	0,2	0,0	0,1	0,4	0,1
0,20	0,1	0,0	0,4	0,0	0,0	0,0	0,2	0,0
0,25	0,0	-0,3	0,2	-0,3	-0,1	0,0	0,2	0,0
0,30	0,1	-0,1	0,3	-0,1	0,0	0,0	0,2	0,0
0,50	0,1	0,0	0,2	0,0	0,0	-0,1	0,2	0,0
0,70	0,0	0,0	0,1	0,0	-0,1	-0,1	0,0	0,0
0,99	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Table 2 – Bias (p.p) calculated on 1000 sample estimates for the growth rate g. Simulation 1: ε ~ N(0, 0.25) , $cor(Y^t, Y^{t-4}) = \mathbf{0.92}$

overlap	**rho=0**				**rho=0.5**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	0.5	-0.5	0.5	-0.5	0.0	0.1	0.2	0.1
0.10	0.1	-0.1	0.0	-0.1	0.1	-0.5	0.2	-0.5
0.15	0.3	-0.2	0.3	-0.2	0.2	-0.1	0.2	-0.1
0.20	0.2	0.2	0.2	0.2	0.0	0.0	0.0	0.0
0.25	-0.1	-0.1	-0.1	-0.1	0.2	0.0	0.4	0.0
0.30	0.0	-0.1	0.0	-0.1	0.2	-0.1	0.1	-0.1
0.50	-0.3	-0.1	-0.3	-0.1	0.2	0.0	0.3	0.0
0.70	0.0	-0.1	0.0	-0.1	0.1	0.0	0.1	0.0
0.99	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.1
overlap	**rho=0.6**				**rho=0.7**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	-0.2	-0.1	-0.2	0.0	-0.4	-0.1	-0.5	-0.1
0.10	0.2	0.0	0.2	0.0	0.3	0.0	0.5	0.0
0.15	-0.2	-0.3	-0.2	-0.3	0.0	0.0	0.0	0.0
0.20	0.5	0.0	0.6	0.0	0.0	0.1	0.0	0.1
0.25	0.1	0.0	0.1	0.0	-0.1	-0.1	-0.1	-0.1
0.30	0.0	-0.2	-0.1	-0.2	0.1	0.0	0.2	0.0
0.50	0.2	-0.1	0.3	-0.1	0.4	0.1	0.5	0.1
0.70	0.0	-0.1	0.1	-0.1	0.0	0.0	0.0	0.0
0.99	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	-0.1	-0.1
overlap	**rho=0.8**				**rho=0.9**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	0.0	-0.8	0.1	-0.8	0.0	-0.1	0.1	0.0
0.10	0.3	-0.3	0.3	-0.3	-0.1	-0.1	0.0	-0.1
0.15	-0.2	-0.1	-0.2	-0.1	0.2	0.0	0.3	0.1
0.20	0.1	0.0	0.1	0.0	-0.1	-0.2	-0.1	-0.1
0.25	0.3	0.1	0.4	0.1	-0.1	-0.1	-0.1	-0.1
0.30	0.2	-0.2	0.4	-0.1	0.3	0.2	0.5	0.2
0.50	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
0.70	-0.1	-0.1	-0.1	0.0	0.0	0.0	0.1	0.1
0.99	0.0	0.0	0.0	0.0	-0.1	-0.1	-0.1	-0.1
overlap	**rho=0.95**				**rho=1**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	-0.1	0.1	0.1	0.2	0.5	-0.5	0.5	-0.5
0.10	-0.1	-0.2	0.0	-0.1	0.1	-0.1	0.0	-0.1
0.15	-0.1	-0.2	0.0	-0.1	0.3	-0.2	0.3	-0.2
0.20	-0.1	-0.1	0.1	-0.1	0.2	0.2	0.2	0.2
0.25	-0.1	0.0	0.0	0.0	-0.1	-0.1	-0.1	-0.1
0.30	0.0	0.1	0.3	0.1	0.0	-0.1	0.0	-0.1
0.50	-0.2	-0.1	0.0	-0.1	-0.3	-0.1	-0.3	-0.1
0.70	-0.2	-0.1	-0.1	-0.1	0.0	-0.1	0.0	-0.1
0.99	-0.1	-0.1	-0.1	-0.1	0.0	0.0	0.0	0.0

Table 3 – Bias (p.p) calculated on 1000 sample estimates for the growth rate g. Simulation 3: ε ~ N(0, 0.35) , $\boldsymbol{cor(Y^t, Y^{t-4}) = 0.86}$

overlap	rho=0				rho=0.5			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	0.5	0.3	0.5	0.3	0.1	0.4	0.1	0.4
0.10	0.4	0.4	0.3	0.4	0.0	0.1	0.1	0.1
0.15	0.1	0.2	0.1	0.2	0.1	0.1	0.1	0.1
0.20	0.3	-0.2	0.3	-0.2	0.4	-0.1	0.4	-0.1
0.25	0.3	0.0	0.2	0.0	0.0	0.1	0.0	0.1
0.30	0.2	0.0	0.2	0.0	0.2	0.1	0.2	0.1
0.50	0.1	-0.1	0.1	-0.1	0.3	0.2	0.3	0.2
0.70	0.2	0.1	0.2	0.1	0.0	-0.1	0.0	-0.1
0.99	0.0	0.0	0.0	0.0	-0.1	-0.1	-0.1	-0.1

overlap	rho=0.6				rho=0.7			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	0.0	0.7	0.1	0.7	0.2	0.5	0.2	0.6
0.10	0.3	-0.2	0.4	-0.2	0.5	-0.2	0.6	-0.2
0.15	0.4	0.2	0.4	0.2	-0.1	0.0	0.0	0.1
0.20	0.0	-0.2	0.1	-0.1	0.1	0.1	0.2	0.1
0.25	0.1	-0.1	0.1	-0.1	-0.1	-0.3	-0.1	-0.3
0.30	0.2	0.0	0.1	0.0	0.1	-0.3	0.3	-0.2
0.50	0.0	-0.1	0.0	0.0	0.1	0.0	0.2	0.0
0.70	0.2	0.1	0.2	0.1	0.1	0.0	0.1	0.0
0.99	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0

overlap	rho=0.8				rho=0.9			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	0.2	-0.2	0.2	-0.1	-0.1	-0.2	0.1	-0.1
0.10	0.2	-0.1	0.3	0.0	0.1	-0.1	0.3	-0.1
0.15	0.4	0.0	0.5	0.1	0.0	0.1	0.1	0.2
0.20	0.1	0.1	0.1	0.2	0.2	-0.2	0.4	-0.1
0.25	0.0	0.3	0.1	0.4	0.2	0.2	0.3	0.3
0.30	0.3	0.0	0.6	0.0	0.0	-0.1	0.1	-0.1
0.50	0.1	0.0	0.3	0.0	-0.1	-0.1	0.0	-0.1
0.70	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0
0.99	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

overlap	rho=0.95				rho=1			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	0.1	0.1	0.5	0.1	0.0	0.4	0.3	0.3
0.10	0.0	-0.2	-0.1	0.0	0.0	0.2	0.3	0.2
0.15	0.2	-0.3	0.5	-0.2	0.0	0.0	0.5	0.2
0.20	0.0	-0.1	0.2	0.0	0.0	0.0	0.0	0.1
0.25	0.3	-0.1	0.6	-0.1	-0.1	-0.1	0.2	0.0
0.30	0.0	0.0	-0.1	0.0	0.0	0.2	0.2	0.2
0.50	0.0	0.0	0.2	0.0	0.2	0.3	0.3	0.3
0.70	0.0	0.0	0.1	0.0	-0.1	-0.1	-0.3	-0.1
0.99	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1

Table 4 – Mean squared error calculated on 1000 sample estimates for the growth rate g. Simulation 1: ε ~ N(0, 0.15) , $cor(Y^t, Y^{t-4}) = \mathbf{0.97}$

overlap	rho=0				rho=0.5			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	45.1	23.1	43.8	23.1	36.4	22.1	46.6	22.1
0.10	42.3	12.3	39.7	12.3	34.8	13.0	43.6	13.0
0.15	41.1	8.4	38.5	8.4	30.3	7.8	38.5	7.8
0.20	39.7	6.3	37.3	6.3	30.3	6.8	37.5	6.8
0.25	34.8	5.8	32.5	5.8	25.0	4.8	33.7	4.8
0.30	36.0	4.8	33.6	4.8	25.1	4.4	31.5	4.4
0.50	25.0	2.6	24.0	2.6	18.5	2.6	23.1	2.6
0.70	15.2	2.0	14.5	2.0	12.3	2.0	15.2	2.0
0.99	1.7	1.2	1.7	1.2	1.7	1.4	1.7	1.4
o	0.03		0.08		0.04		0.08	
overlap	rho=0.6				rho=0.7			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	30.3	24.0	42.3	24.0	28.1	23.1	45.0	23.1
0.10	29.4	13.0	43.9	13.0	24.0	11.6	37.2	12.3
0.15	28.1	8.4	41.1	8.4	23.0	9.0	38.5	9.0
0.20	26.0	6.3	39.7	6.3	22.1	7.3	37.2	7.3
0.25	22.1	5.3	31.4	5.3	21.2	4.8	33.7	4.8
0.30	21.3	4.4	30.5	4.4	19.4	4.4	31.5	4.4
0.50	16.0	2.3	23.0	2.3	15.2	2.6	25.0	2.6
0.70	10.9	1.7	16.0	1.7	9.6	2.0	14.5	2.0
0.99	1.7	1.4	1.7	1.4	1.7	1.2	2.0	1.2
o	0.05		0.08		0.06		0.08	
overlap	rho=0.8				rho=0.9			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	23.1	24.1	46.3	24.1	14.4	24.0	46.3	24.0
0.10	21.3	12.3	42.4	13.1	12.3	13.0	43.6	13.0
0.15	19.4	8.4	39.8	8.4	10.9	9.0	37.3	9.0
0.20	18.5	6.3	37.2	6.3	10.9	6.3	34.9	6.8
0.25	16.8	4.8	33.7	4.8	10.3	4.9	33.7	4.8
0.30	16.0	4.0	33.8	4.0	10.9	4.4	32.6	4.4
0.50	11.6	2.3	22.1	2.6	7.3	2.6	22.3	2.6
0.70	7.3	1.7	14.4	1.7	5.3	2.0	14.5	2.0
0.99	1.5	1.2	1.7	1.2	1.2	1.2	1.7	1.2
o	0.08		0.08		0.15		0.08	
overlap	rho=0.95				rho=1			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	8.4	24.1	44.9	25.0	1.2	22.2	43.6	23.1
0.10	8.4	11.6	41.0	12.3	1.2	11.6	39.7	12.3
0.15	6.8	8.5	34.8	9.0	1.2	9.0	36.2	9.0
0.20	7.3	6.3	37.4	6.3	1.2	6.3	37.3	6.3
0.25	6.3	5.4	32.5	5.4	1.2	4.8	33.7	4.8
0.30	6.3	4.0	36.1	4.4	1.2	4.4	31.4	4.4
0.50	5.8	2.6	26.1	2.6	1.2	2.6	23.1	2.9
0.70	3.2	2.0	14.5	2.0	1.2	1.7	15.2	1.7
0.99	1.2	1.2	1.7	1.2	1.2	1.2	2.0	1.2
o	0.29		0.08		1.0		0.08	

Table 5 – Mean squared error calculated on 1000 sample estimates for the growth rate g. Simulation 2: ε ~ N(0, 0.25) , $\boldsymbol{cor(Y^t, Y^{t-4}) = 0.92}$

overlap	rho=0				rho=0.5			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	52.1	67.5	49.3	67.5	34.8	64.0	43.6	64.0
0.10	44.9	34.8	41.0	34.8	34.8	36.3	42.3	36.3
0.15	41.1	22.1	38.5	22.1	33.7	23.1	41.0	23.1
0.20	39.7	18.5	37.3	18.5	30.3	17.6	37.2	17.6
0.25	37.2	13.7	34.8	13.7	27.1	14.4	33.8	14.4
0.30	36.0	11.6	33.6	11.6	26.1	12.3	31.4	12.3
0.50	27.1	7.3	26.1	7.3	21.2	6.8	27.1	6.8
0.70	18.5	5.3	16.8	5.3	14.5	4.8	17.7	4.8
0.99	4.0	3.6	4.0	3.6	4.0	3.6	4.0	3.6
o	0.09		0.09		0.11		0.09	
overlap	**rho=0.6**				**rho=0.7**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	31.4	65.6	44.9	65.6	26.2	67.3	41.2	67.3
0.10	31.4	33.6	43.6	33.6	28.2	33.6	43.8	34.8
0.15	30.3	24.1	42.3	24.1	27.0	21.2	42.3	22.1
0.20	24.3	17.6	34.0	17.6	24.0	17.7	38.4	17.7
0.25	27.1	14.4	37.2	15.2	23.1	15.2	36.0	15.2
0.30	26.0	11.6	34.8	12.3	23.1	13.0	36.0	13.0
0.50	19.4	7.3	27.1	7.3	16.2	7.3	24.3	7.3
0.70	13.7	4.9	17.7	4.9	10.9	5.3	16.8	5.3
0.99	3.6	3.6	4.0	3.6	4.0	4.0	4.4	4.0
o	0.13		0.09		0.16		0.09	
overlap	**rho=0.8**				**rho=0.9**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	22.1	67.9	42.3	71.2	16.0	60.9	46.3	64.0
0.10	22.2	37.3	45.0	37.3	16.0	32.5	43.6	33.7
0.15	20.3	24.0	39.7	24.0	15.3	20.3	41.1	21.2
0.20	19.4	16.8	38.5	16.8	14.5	18.5	38.5	19.4
0.25	21.3	14.5	41.1	15.2	13.0	13.7	36.0	13.7
0.30	16.0	11.6	31.5	11.6	11.7	11.6	33.9	11.6
0.50	13.0	7.3	24.0	7.3	10.3	7.3	27.1	7.9
0.70	9.6	4.9	16.8	4.8	7.8	4.8	17.7	5.3
0.99	3.6	3.6	4.0	3.6	3.6	3.6	4.0	3.6
o	0.22		0.09		0.40		0.09	
overlap	**rho=0.95**				**rho=1**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	9.0	62.4	42.3	65.7	3.5	61.1	49.3	65.9
0.10	9.0	36.0	43.6	38.5	3.3	33.7	41.0	36.0
0.15	9.6	22.1	43.6	24.0	3.3	21.2	45.0	23.1
0.20	9.0	17.7	37.2	18.5	3.3	16.0	38.5	18.5
0.25	8.4	14.4	38.4	15.2	3.3	13.7	37.2	15.2
0.30	7.8	11.6	34.9	12.3	3.2	11.6	32.5	13.0
0.50	6.3	6.8	25.0	7.3	3.3	6.8	25.1	7.3
0.70	5.3	5.3	16.0	5.3	3.2	4.9	16.0	5.3
0.99	2.9	2.9	3.6	3.3	2.9	3.2	4.0	3.6
o	0.79		0.09		1.00		0.09	

Table 6 – Mean squared error calculated on 1000 sample estimates for the growth rate g. Simulation 3: ε ~ N(0, 0.35) , $cor(Y^t, Y^{t-4}) = \mathbf{0.86}$

overlap	rho=0				rho=0.5			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	52.1	132.3	47.9	151.4	37.2	130.1	47.6	151.5
0.10	50.6	74.1	46.3	75.9	37.2	67.3	46.3	75.7
0.15	46.3	50.5	43.6	50.5	34.8	47.6	43.6	50.4
0.20	45.0	39.7	41.1	37.3	36.2	37.2	41.1	37.2
0.25	39.8	30.3	38.5	29.2	33.6	29.2	38.4	29.2
0.30	39.7	23.0	37.3	25.0	31.4	23.1	37.3	25.0
0.50	29.2	14.5	28.1	14.5	24.1	14.5	28.2	14.5
0.70	20.3	10.3	19.4	10.3	17.6	10.3	19.4	10.3
0.99	7.8	7.3	7.8	7.3	7.3	7.3	7.9	7.3
o	0.17		0.17		0.22		0.17	
overlap	**rho=0.6**				**rho=0.7**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	34.8	132.7	47.6	151.8	33.7	127.9	47.7	151.7
0.10	33.7	70.6	46.4	75.7	33.9	70.6	46.6	75.7
0.15	33.8	46.3	43.7	50.5	29.2	43.6	43.6	50.4
0.20	32.5	37.3	41.0	37.2	27.1	32.5	41.0	37.2
0.25	29.2	28.1	38.5	29.2	25.0	29.3	38.5	29.3
0.30	30.3	25.0	37.2	25.0	26.0	23.1	37.3	25.0
0.50	22.1	14.5	28.1	14.4	18.5	15.2	28.1	14.4
0.70	14.5	9.0	19.4	10.3	14.5	10.2	19.4	10.2
0.99	6.8	6.8	7.9	7.3	7.8	7.3	7.8	7.3
o	0.25		0.17		0.30		0.17	
overlap	**rho=0.8**				**rho=0.9**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	27.1	139.3	47.7	151.3	20.3	127.7	47.6	151.3
0.10	23.1	64.0	46.3	75.7	19.4	67.3	46.3	75.7
0.15	23.2	47.6	43.8	50.4	16.8	44.9	43.6	50.5
0.20	24.0	34.8	41.0	37.3	17.7	33.7	41.1	37.2
0.25	23.0	26.1	38.5	29.3	16.0	29.2	38.5	29.3
0.30	21.3	24.0	37.6	25.0	14.4	22.1	37.2	25.0
0.50	18.5	14.4	28.2	14.4	13.0	14.5	28.1	14.5
0.70	13.0	10.9	19.4	10.2	10.2	10.2	19.4	10.2
0.99	6.3	6.3	7.8	7.3	5.8	5.8	7.8	7.3
o	0.41		0.17		0.74		0.17	
overlap	**rho=0.95**				**rho=1**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	13.7	123.2	47.9	151.3	5.8	116.8	47.7	151.4
0.10	12.3	62.5	46.3	75.7	5.8	64.0	46.3	75.7
0.15	13.0	45.0	43.8	50.5	5.8	37.2	43.8	50.5
0.20	12.3	33.7	41.0	37.2	5.8	29.2	41.0	37.2
0.25	11.0	26.0	38.8	29.2	6.3	22.1	38.5	29.2
0.30	13.0	23.0	37.2	25.0	6.3	22.1	37.3	25.0
0.50	10.9	13.0	28.1	14.4	6.3	11.7	28.2	14.5
0.70	9.0	9.6	19.4	10.2	6.3	8.4	19.5	10.3
0.99	6.3	6.3	7.8	7.3	5.8	5.8	7.9	7.3
o	1.0		0.17		1.0		0.17	

Table 7 – Percentage of times that the confidence interval of the estimates contains the true value of the population. Simulation 1: $\varepsilon \sim N(0, 0.15)$, $\boldsymbol{cor(Y^t, Y^{t-4}) = 0.97}$

overlap	rho=0				rho=0.5			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	95.5	92.5	95.6	92.5	95.3	90.5	95.0	90.5
0.10	96.1	93.5	95.1	93.6	93.8	91.4	93.1	91.4
0.15	95.3	94.1	95.9	94.2	94.1	94.4	94.3	94.5
0.20	94.0	94.1	93.9	94.1	93.9	93.4	94.9	93.4
0.25	93.8	92.8	94.6	92.9	95.2	93.7	94.5	93.7
0.30	94.1	91.3	93.3	91.4	95.6	94.9	95.3	94.9
0.50	95.6	93.3	94.6	93.3	94.9	95.2	94.2	95.2
0.70	94.1	94.9	94.1	94.8	95.1	94.9	95.2	94.9
0.99	96.1	96.1	95.9	96.1	94.1	93.9	95.0	93.9
overlap	**rho=0.6**				**rho=0.7**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	94.6	88.5	95.2	88.7	95.4	90.3	95.4	90.6
0.10	94.8	92.6	94.7	93.1	95.0	94.1	94.5	94.4
0.15	94.3	93.4	93.2	93.6	95.7	91.6	95.6	91.7
0.20	94.6	93.9	92.9	93.9	95.6	92.9	94.8	93.2
0.25	96.0	93.6	96.9	93.6	94.9	94.7	94.9	94.9
0.30	95.3	92.7	95.7	92.7	95.2	93.9	95.5	93.8
0.50	94.1	95.2	94.0	95.4	94.0	93.8	93.4	94.0
0.70	94.7	94.4	93.4	94.5	95.0	93.2	95.5	93.4
0.99	95.3	94.6	94.6	94.6	93.1	95.4	94.4	95.4
overlap	**rho=0.8**				**rho=0.9**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	93.4	90.2	93.5	90.8	94.3	90.0	93.4	90.5
0.10	94.8	92.3	95.2	92.5	95.5	92.4	94.4	92.6
0.15	95.6	92.8	95.2	93.0	95.6	92.6	95.1	93.1
0.20	94.9	93.9	95.6	94.2	95.3	92.8	95.0	93.2
0.25	94.9	95.0	94.7	95.3	95.1	93.6	94.6	93.7
0.30	95.1	93.5	94.6	93.3	95.1	93.1	94.9	93.3
0.50	94.8	94.7	95.3	94.7	95.8	94.0	95.6	94.3
0.70	95.9	94.9	95.3	95.1	94.3	94.4	96.0	94.3
0.99	95.6	95.0	95.0	95.2	95.6	95.0	95.4	95.0
overlap	**rho=0.95**				**rho=1**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	94.9	90.2	94.9	90.6	95.6	89.8	95.4	91.1
0.10	95.1	92.8	94.7	92.9	94.3	92.2	95.0	93.0
0.15	95.8	93.0	96.7	93.8	94.4	90.7	95.9	91.4
0.20	94.2	92.2	94.7	92.4	95.1	93.2	94.9	93.9
0.25	96.0	93.5	96.0	93.3	94.1	94.1	94.6	94.7
0.30	94.9	94.4	94.8	94.9	93.5	93.0	95.5	93.3
0.50	94.3	94.7	93.6	95.0	95.0	92.9	95.0	92.9
0.70	94.7	94.3	95.7	94.1	94.9	94.1	94.0	95.1
0.99	94.5	94.7	95.1	95.1	94.0	94.0	93.8	94.3

Table 8 – Percentage of times that the confidence interval of the estimates contains the true value of the population. Simulation 2: ε ~ N(0, 0.25) , $\boldsymbol{cor(Y^t, Y^{t-4}) = 0.92}$

overlap	**rho=0**				**rho=0.5**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	94.4	91.1	94.2	91.4	95.0	90.9	95.6	91.2
0.10	95.6	91.4	95.8	91.7	94.9	92.2	95.5	92.7
0.15	95.0	94.3	95.1	94.4	94.8	94.9	95.0	94.9
0.20	95.6	92.9	95.6	93.0	95.5	94.6	94.6	94.7
0.25	94.6	95.3	95.7	95.3	95.4	93.4	96.0	93.4
0.30	95.6	94.0	95.7	94.0	94.9	94.4	95.5	94.4
0.50	94.3	93.8	94.5	93.8	94.6	95.1	94.1	95.1
0.70	95.1	94.4	95.1	94.4	95.4	95.3	94.5	95.3
0.99	95.1	94.0	95.3	94.0	95.5	94.7	95.0	94.8
overlap	**rho=0.6**				**rho=0.7**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	94.8	91.4	94.9	92.2	95.7	91.7	95.1	91.8
0.10	94.7	93.0	94.7	93.0	94.1	92.9	94.7	93.1
0.15	95.3	93.1	94.5	93.3	94.2	95.2	94.3	95.6
0.20	96.6	94.2	95.7	94.2	94.6	93.6	94.9	94.0
0.25	93.9	93.2	94.1	93.3	94.7	94.3	95.3	94.5
0.30	94.1	94.6	94.2	94.7	94.2	92.9	93.7	92.8
0.50	94.4	94.4	94.0	94.4	95.4	94.2	95.8	94.3
0.70	93.6	94.8	94.5	95.0	95.3	94.1	94.6	93.9
0.99	95.6	95.3	95.6	95.6	93.9	94.4	94.0	94.3
overlap	**rho=0.8**				**rho=0.9**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	96.0	90.3	95.9	91.2	95.2	90.8	95.0	91.2
0.10	94.6	92.1	94.0	92.5	93.9	93.1	94.4	93.7
0.15	95.0	92.7	95.1	93.2	94.4	94.9	94.1	95.2
0.20	95.5	94.3	95.0	95.0	94.9	92.6	94.3	93.0
0.25	93.3	91.6	93.7	91.6	94.6	94.7	94.0	94.8
0.30	96.1	94.7	96.0	95.0	95.3	93.5	95.1	93.6
0.50	94.8	94.2	95.1	94.2	93.9	93.0	92.8	93.1
0.70	95.2	93.6	94.8	93.8	94.8	94.2	94.7	94.2
0.99	93.9	94.1	94.1	93.8	94.7	94.2	95.3	94.2
overlap	**rho=0.95**				**rho=1**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	95.7	92.3	95.7	92.6	94.7	88.5	94.6	90.0
0.10	94.4	90.9	94.3	91.4	94.7	90.7	95.5	92.0
0.15	93.8	92.7	93.3	92.8	94.9	93.5	93.7	94.2
0.20	94.3	93.0	95.1	93.6	94.3	92.6	94.8	93.5
0.25	95.3	92.8	94.5	93.2	94.9	92.5	94.3	92.8
0.30	94.7	94.2	95.3	94.2	94.7	93.1	95.3	93.6
0.50	95.0	94.6	94.6	94.6	94.9	93.4	94.8	93.9
0.70	95.1	93.0	94.8	93.7	93.5	94.4	95.9	94.5
0.99	96.0	95.9	96.6	96.4	96.2	95.7	95.2	95.5

Table 9 – Percentage of times that the confidence interval of the estimates contains the true value of the population. Simulation 3: ε ~ N(0, 0.35) , $\boldsymbol{cor(Y^t, Y^{t-4}) = 0.86}$

overlap	rho=0				rho=0.5			
	calibration		no calibration		calibration		no calibration	
	Gall.cal	Golp.cal	Gall	Golp	Gall.cal	Golp.cal	Gall	Golp
0.05	94.3	91.5	94.8	91.7	95.1	91.4	95.1	91.5
0.10	95.2	92.4	94.2	92.6	93.4	93.3	94.6	93.2
0.15	94.7	92.6	94.9	92.5	95.2	91.6	94.9	91.7
0.20	94.5	92.5	94.3	92.5	94.4	92.6	94.9	92.6
0.25	95.8	93.6	95	93.6	94	93.5	93.7	93.4
0.30	95.2	95.4	94.3	95.4	94.4	94.7	93.7	94.6
0.50	94.4	94.3	94.6	94.3	94.5	94	94.6	94.2
0.70	94.8	94.1	94.5	94.1	95.1	93.8	94.8	93.9
0.99	94.1	94.8	94.1	95	95	95.4	94.9	95.2
overlap	**rho=0.6**				**rho=0.7**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	95.5	91.4	95.5	91.6	94.9	91.2	94.4	91.3
0.10	95.2	92.8	94.6	93	93.7	91.7	94.4	91.7
0.15	94.4	92.9	94.4	93.1	94.3	94.3	94.6	94.5
0.20	95	93.2	94.9	93.1	96	94.8	96	95.1
0.25	95.1	93.9	94.4	93.6	95.1	93.5	94.5	93.8
0.30	93.4	94.3	92.9	94.2	94.9	94	94.4	94.1
0.50	94.5	93.6	95.5	93.6	96.5	93	96	93
0.70	95.3	95.2	95.6	95.2	95.3	95.3	94.7	95.5
0.99	95.5	94.6	95.5	94.9	94.4	94.1	94.7	94
overlap	**rho=0.8**				**rho=0.9**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	94.8	90.8	94.7	90.7	94.4	90.8	93.8	91.5
0.10	95	92.4	95.2	93	94.3	91.2	94.8	91.7
0.15	95.7	92.4	94.6	92.9	96	93.2	95.5	93.8
0.20	94.9	92.6	93.9	92.6	93.4	93.8	94.1	93.9
0.25	93.7	94.3	93.2	94.5	95	92.8	94.9	93.4
0.30	94.8	93.8	94.8	94.4	95.6	93.9	95.4	93.8
0.50	94.1	94.6	94.7	94.6	95	94.1	95.5	94.6
0.70	95.2	92.7	95.1	92.8	94	93.3	94.3	93.5
0.99	94.4	95.3	94.7	95.2	96.7	96.3	96.5	96.7
overlap	**rho=0.95**				**rho=1**			
	calibration		**no calibration**		**calibration**		**no calibration**	
	Gall.cal	**Golp.cal**	**Gall**	**Golp**	**Gall.cal**	**Golp.cal**	**Gall**	**Golp**
0.05	94.6	90.8	94.3	91.7	95.2	88.5	94.6	91
0.10	95.1	93.5	95.9	93.7	95.3	91.1	94.9	93
0.15	94.9	93.1	93.7	93.6	93.9	92.6	94.9	93.5
0.20	94.6	93	95.2	93.5	95.1	93.9	95.3	94.2
0.25	95.6	95.2	95.3	95.2	92.7	93.6	94.3	94.7
0.30	92.4	93.5	93.8	93.8	92.8	92.5	93.9	93.2
0.50	93.5	94.9	95.3	95.1	94.2	94.5	94.6	94.8
0.70	94.7	95.2	94.8	95.3	94.5	94.7	94.4	94.5
0.99	95.4	95.5	95.5	95.3	96.3	95.4	95.1	95.2

Table 10 – Coefficients of variation calculated on 1000 sample estimates for the total estimation of Y^{t-4}, using calibration. Simulation 1: ε ~ N(0, 0.15)

overlap	rho=0		rho=0.5		rho=0.6		rho=0.7	
	Gall.cal	Golp.cal	Gall.cal	Golp.cal	Gall.cal	Golp.cal	Gall.cal	Golp.cal
0.05	5.1	23.1	4.5	20.1	4.4	19.2	4.0	18.0
0.10	5.2	17.0	4.6	14.2	4.4	14.0	3.9	13.1
0.15	5.2	14.1	4.7	12.2	4.4	11.0	4.0	10.9
0.20	5.2	12.1	4.5	9.9	4.3	9.4	4.0	9.2
0.25	5.1	10.7	4.5	9.2	4.2	8.6	4.1	8.5
0.30	5.3	9.7	4.6	8.7	4.4	8.0	4.0	7.4
0.50	5.3	7.9	4.7	6.6	4.4	6.0	4.2	5.7
0.70	5.4	6.3	4.7	5.4	4.2	5.1	3.9	4.7
0.99	5.4	5.4	4.6	4.6	4.2	4.2	4.1	4.1
overlap	**rho=0.8**		**rho=0.9**		**rho=0.95**		**rho=1**	
	Gall.cal	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**
0.05	3.8	16.5	2.9	12.8	2.3	10.3	1.2	5.5
0.10	3.7	11.6	3.0	9.2	2.4	7.5	1.3	3.9
0.15	3.5	9.4	2.9	7.9	2.4	6.1	1.3	3.4
0.20	3.6	8.1	3.0	6.6	2.4	5.6	1.2	2.8
0.25	3.7	7.5	2.8	5.9	2.3	4.7	1.3	2.5
0.30	3.6	6.4	3.1	5.1	2.3	4.3	1.3	2.3
0.50	3.6	5.2	2.9	4.1	2.4	3.3	1.2	1.8
0.70	3.6	4.3	3.0	3.5	2.3	2.9	1.2	1.5
0.99	3.8	3.8	2.9	2.9	2.2	2.3	1.2	1.3

Table 11 – Coefficients of variation calculated on 1000 sample estimates for the total estimation of Y^{t}, using calibration. Simulation 1: ε ~ N(0, 0.15)

overlap	rho=0		rho=0.5		rho=0.6		rho=0.7	
	Gall.cal	Golp.cal	Gall.cal	Golp.cal	Gall.cal	Golp.cal	Gall.cal	Golp.cal
0.05	5.2	23.9	4.8	20.3	4.3	19.6	4.2	17.9
0.10	5.5	17.7	4.8	14.8	4.2	14.4	4.2	13.4
0.15	5.5	14.6	4.7	12.4	4.5	11.3	4.3	11.0
0.20	5.3	12.6	4.7	10.3	4.3	9.7	4.2	9.4
0.25	5.6	11.2	4.7	9.5	4.2	8.9	4.2	8.6
0.30	5.5	10	4.7	8.9	4.2	8.3	4.1	7.7
0.50	5.6	8.2	4.8	6.8	4.3	6.2	4.2	5.9
0.70	5.4	6.5	4.6	5.6	4.4	5.3	4.1	4.8
0.99	5.6	5.6	4.7	4.7	4.3	4.3	4.2	4.2
overlap	**rho=0.8**		**rho=0.9**		**rho=0.95**		**rho=1**	
	Gall.cal	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**
0.05	3.7	16.8	3.0	12.8	2.3	10.0	0.1	0.7
0.10	3.7	12.0	2.8	9.1	2.3	7.1	0.1	0.5
0.15	3.8	9.6	2.8	7.8	2.2	5.9	0.1	0.4
0.20	3.6	8.2	2.8	6.5	2.2	5.3	0.1	0.3
0.25	3.7	7.6	2.9	5.8	2.3	4.5	0.1	0.3
0.30	3.7	6.5	2.8	5.1	2.3	4.1	0.1	0.2
0.50	3.7	5.2	2.9	4.1	2.2	3.1	0.1	0.2
0.70	3.6	4.4	2.8	3.5	2.3	2.8	0.1	0.1
0.99	3.8	3.8	2.9	2.9	2.1	2.1	0.1	0.1

Table 12 – Coefficients of variation calculated on 1000 sample estimates for the total estimation of Y^{t-4}, using calibration. Simulation 2: ε ~ N(0, 0.25)

overlap	**rho=0**		**rho=0.5**		**rho=0.6**		**rho=0.7**	
	Gall.cal	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**
0.05	5.4	23.9	4.6	20.2	4.3	19.2	4.2	18.0
0.10	5.3	16.5	4.6	14.8	4.5	14.0	4.1	13.4
0.15	5.1	13.8	4.6	12.3	4.3	11.3	4.1	11.2
0.20	5.1	11.5	4.6	10.5	4.4	10.1	4.1	9.8
0.25	5.3	10.5	4.5	9.4	4.4	8.8	4.3	8.5
0.30	5.3	9.8	4.7	8.8	4.5	8.3	4.3	7.8
0.50	5.5	7.5	4.7	6.9	4.4	6.2	4.1	6.0
0.70	5.3	6.5	4.6	5.4	4.3	5.2	4.2	5.1
0.99	5.4	5.5	4.7	4.8	4.5	4.5	4.1	4.2
overlap	**rho=0.8**		**rho=0.9**		**rho=0.95**		**rho=1**	
	Gall.cal	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**
0.05	3.6	17.1	3.2	15.0	2.6	12.5	2.0	9.1
0.10	3.7	12.3	3.2	10.3	2.6	8.5	2.1	6.6
0.15	3.8	9.7	3.2	8.4	2.7	6.6	2.0	5.2
0.20	3.8	8.7	3.1	7.1	2.7	6.0	2.0	4.6
0.25	3.8	7.3	3.3	6.5	2.5	5.1	2.0	4.2
0.30	3.7	6.9	3.1	5.9	2.7	4.8	2.0	3.8
0.50	3.7	5.2	3.2	4.6	2.5	3.6	2.0	2.9
0.70	3.6	4.4	3.2	3.8	2.7	3.2	2.0	2.4
0.99	3.7	3.7	3.3	3.3	2.5	2.5	1.9	2.0

Table 13 – Coefficients of variation calculated on 1000 sample estimates for the total estimation of Y^{t}, using calibration. Simulation 2: ε ~ N(0, 0.25)

overlap	**rho=0**		**rho=0.5**		**rho=0.6**		**rho=0.7**	
	Gall.cal	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**
0.05	5.8	25.1	5.0	20.9	4.7	20.7	4.2	18.4
0.10	5.6	17.9	5.1	15.7	4.6	15.0	4.4	13.6
0.15	5.6	14.3	4.9	12.6	4.6	12.2	4.4	11.4
0.20	5.5	12.3	4.9	11.2	4.4	10.8	4.3	9.9
0.25	5.7	11.2	4.8	10.0	4.6	9.4	4.4	8.8
0.30	5.5	10.3	4.8	9.1	4.7	8.6	4.4	8.0
0.50	5.6	8.1	5.0	7.2	4.7	6.6	4.4	6.3
0.70	5.9	6.9	4.9	5.8	4.6	5.4	4.4	5.4
0.99	5.9	5.9	5.0	5.0	4.7	4.7	4.3	4.4
overlap	**rho=0.8**		**rho=0.9**		**rho=0.95**		**rho=1**	
	Gall.cal	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**
0.05	3.8	16.8	3.2	14.1	2.2	10.7	0.0	0.0
0.10	3.8	12.7	3.2	9.8	2.3	7.1	0.0	0.0
0.15	3.8	10.1	3.2	8.1	2.3	5.8	0.0	0.0
0.20	3.8	8.9	3.3	7.0	2.2	5.0	0.0	0.0
0.25	4.0	7.4	3.1	6.4	2.2	4.5	0.0	0.0
0.30	3.8	7.1	3.1	5.8	2.2	4.2	0.0	0.0
0.50	3.7	5.4	3.2	4.4	2.2	3.1	0.0	0.0
0.70	3.7	4.4	3.2	3.8	2.1	2.6	0.0	0.0
0.99	3.7	3.7	3.2	3.2	2.1	2.1	0.0	0.0

Table 14 – Coefficients of variation calculated on 1000 sample estimates for the total estimation of Y^{t-4}, using calibration. Simulation 3: ε ~ N(0, 0.35)

overlap	rho=0		rho=0.5		rho=0.6		rho=0.7	
	Gall.cal	Golp.cal	Gall.cal	Golp.cal	Gall.cal	Golp.cal	Gall.cal	Golp.cal
0.05	5.5	24.0	4.5	19.6	4.2	19.6	4.2	19.5
0.10	5.2	17.3	4.4	14.5	4.5	14.3	4.4	13.7
0.15	5.3	13.9	4.8	11.9	4.4	11.5	4.3	10.6
0.20	5.2	12.1	4.7	10.7	4.6	10.3	4.1	9.7
0.25	5.2	10.7	4.6	9.6	4.3	9.3	4.2	8.8
0.30	5.2	9.6	4.7	8.6	4.5	8.5	4.2	7.8
0.50	5.0	7.4	4.7	6.7	4.5	6.4	4.1	6.0
0.70	5.1	6.2	4.8	5.6	4.5	5.3	4.4	5.2
0.99	5.1	5.2	4.7	4.7	4.5	4.5	4.5	4.5
overlap	**rho=0.8**		**rho=0.9**		**rho=0.95**		**rho=1**	
	Gall.cal	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**
0.05	4.0	17.4	3.5	15.4	3.0	13.9	2.7	12.6
0.10	3.7	12.5	3.5	10.8	3.0	9.6	2.7	9.2
0.15	3.9	10.2	3.5	9.3	3.1	8.1	2.7	7.0
0.20	4.1	8.7	3.3	7.8	3.0	7.1	2.7	6.1
0.25	3.8	7.8	3.5	7.2	3.0	6.1	2.8	5.3
0.30	3.9	7.2	3.5	6.4	3.2	5.7	2.8	5.2
0.50	4.1	5.7	3.5	5.1	3.2	4.4	2.8	3.8
0.70	3.8	4.7	3.4	4.2	3.1	3.6	2.8	3.3
0.99	4.0	4.1	3.6	3.6	2.9	2.9	2.6	2.7

Table 15– Coefficients of variation calculated on 1000 sample estimates for the total estimation of Y^{t}, using calibration. Simulation 3: ε ~ N(0, 0.35)

overlap	rho=0		rho=0.5		rho=0.6		rho=0.7	
	Gall.cal	Golp.cal	Gall.cal	Golp.cal	Gall.cal	Golp.cal	Gall.cal	Golp.cal
0.05	5.8	26.7	5.2	22.2	4.9	21.8	4.6	20.7
0.10	6.0	19.2	5.2	15.9	5.0	15.7	4.6	14.9
0.15	5.9	15.6	5.2	13.4	5.0	12.4	4.6	12.0
0.20	6.1	13.8	5.4	12.0	5.0	11.1	4.6	10.5
0.25	5.9	12.3	5.3	10.7	5.0	10.2	4.6	9.6
0.30	5.9	10.8	5.2	9.7	5.0	8.9	4.9	8.7
0.50	6.0	8.4	5.1	7.4	5.0	7.0	4.5	6.3
0.70	6.0	7.1	5.2	6.3	4.7	5.6	4.7	5.6
0.99	6.0	6.0	5.2	5.2	5.0	5.1	4.8	4.8
overlap	**rho=0.8**		**rho=0.9**		**rho=0.95**		**rho=1**	
	Gall.cal	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**	**Gall.cal**	**Golp.cal**
0.05	4.0	18.0	3.4	15.0	2.7	12.6	0.0	0.0
0.10	4.2	12.8	3.4	10.7	2.6	8.7	0.0	0.0
0.15	4.0	10.8	3.5	9.2	2.7	7.0	0.0	0.0
0.20	4.0	9.4	3.4	7.9	2.7	6.1	0.0	0.0
0.25	4.3	7.9	3.4	7.0	2.6	5.4	0.0	0.0
0.30	4.2	7.7	3.2	5.9	2.7	4.9	0.0	0.0
0.50	4.1	5.8	3.4	4.9	2.7	3.9	0.0	0.0
0.70	4.1	4.9	3.4	4.1	2.6	3.2	0.0	0.0
0.99	4.2	4.2	3.4	3.5	2.7	2.7	0.0	0.0

References

Andersson C., Andersson K., Lundquist P. (2011), "Estimation of change in a rotation panel design", Proc. 58th World Statistical Congress, Dublin (Session CPS028).

Andreis F., Conti P.L., Marella D., Mecatti F. (2016), "Resampling from finite population under complex designs: the pseudo-population approach", Proceedings 48Th SIS.

Antal E. & Tillé Y. (2012), "A Direct Bootstrap Method for Complex Sampling Designs From a Finite Population", Journal of the American Statistical Association, pp. 534-543.

Bacchini F., Ceccarelli C., Chianella D., Iannaccone R. (2013), "An application of calibration estimators for the quarterly turnover in service sector", Proceedings of the 50th SIEDS scientific meeting. Rome, 31 May 2013.

Bacchini F., Ceccarelli C., Chianella D., Iannaccone R. (2014), "Improving quarterly index of turnover by means of a calibration estimator", Rivista Italiana di Economia Demografia e Statistica, Vol. LXVIII N. 1, pp. 15-22.

Bacchini F., Busanello G., Chianella D., Cinelli R.D., Iannaccone R., Quondamstefano V. (2015), "Recent developments for quarterly service turnover indices", Rivista italiana di statistica ufficiale (numero 1), pp. 21-49.

Barcaroli G., Buglielli D., De Vitiis C. (2010), "MAUSS-R: user and methodological manual", Istituto Nazionale di Statistica.

Berger Y. (2004), "Variance estimation for measures of change in probability sampling", Canadian Journal of Statistics, 32, 4, pp. 451-467.

Berger Y. & Oguz Alper M. (2015), "Variance Estimation of Change in Poverty Rates: an Application to the Turkish EU-SILC Survey", Journal of Official Statistics, Vol. 31(2), pp. 155-175.

Berger Y. & Priam R. (2013), "A simple variance estimator of change for rotating repeated surveys: an application to the EU-SILC survey", paper presented at the NTTS Conference.

Ceccarelli C., Di Marco M., Rinaldelli C. (2008), "L'indagine europea sui redditi e le condizioni di vita delle famiglie (Eu-Silc)", Metodi e Norme n. 37, Istat.

Ceccarelli C., Chianella D., De Filippo F., Graziani C., Guandalini A., Lattanzio M., Loriga S., Martini A., Dionisio Terribili M. (2017), "Quality improvements in variance estimation for the labour force survey", Rivista Italiana di Economia Demografia e Statistica, Vol. LXXI n.4, pp. 15-25.

Ceccarelli C.; Giorgi G. M.; Guandalini A. (2011), "Varianza dello stimatore calibrato in presenza di informazioni ausiliarie campionarie", Rivista Italiana di Economia Demografia e Statistica, Vol. LXV n.1, pp. 53–60.

Chianella D., Cinelli R.D., Quondamstefano V. (2013), "I nuovi indici del fatturato dei servizi: Aspetti metodologici e piano di campionamento", Proceedings of the Istat seminar on short-term indicators of prices for the production of services and turnover of services (available on https://www4.istat.it/it/archivio/90108).

Chianella D., Iannaccone R., Iaconelli B. (2015), "An Estimator for the Growth Rates in Short-Term Business Statistics Using Calibration". Scientific Poster presented at the JOS 30th Anniversary Conference Statistics Sweden: https://www.scb.se/Grupp/Produkter_Tjanster/Kurser/_Dokument/JOS-2015/Accepted-posters.pdf.

Cochran W.G. (1977), "Sampling Techniques", 3rd Edition, John Wiley & Sons, New York.

Conti P., Marella D. (2011), "Campionamento da popolazioni finite: teoria e tecnica", Springer, Milano.

Deville J.C. & Sarndal C.E. (1992), "Calibration estimators in survey sampling", Journal of the American Statistical association, 97, pp. 376-382.

Efron B. (1982), "The jacknife, the bootstrap and other resampling plans", CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38, SIAM, Philadelphia.

EUROSTAT (2006), "Methodology of short-term business statistics: Interpretation and guidelines".

Gazzelloni S. (2006), "La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione", Istituto Nazionale di Statistica, Metodi e norme n. 32.

Guandalini A. (2017), "The variance estimator of differences between estimates of two yearly average of totals", Appunti metodologici.

Goedemé T. (2013), "The EU-SILC sample design variables: critical review and recommendations", CSB Working Paper No. 13 / 02.

Hajek J. (1964), "Asymptotic Theory of Rejective Sampling with Varying Probabilities from a Finite Population", The Annals of Mathematical Statistics, Vol. 35, no. 4, pp. 1491-1523.

Hiridoglou M.A., Särndal C.E., Binder D.A. (1995), "Weighting and estimation in business surveys", pp. 477–502 in: B.G. Cox (ed.), Business Survey Methods, John Wiley &amp; Sons, New York.

Holmberg A. (1998), "A bootstrap approach to probability proportional to size sampling", Proceedings of the survey research methods section of the American Statistical Association, pp. 378-383.

Holt D. & Smith T.M.F. (1979), "Post-stratification", Journal of the Royal Statistical Society, A, 142, 33–46.

Holt D. & Skinner C.J. (1989), "Components of change in repeated surveys", International Statistical Review, 57, pp. 1–18.

Kewski D. & Rao J.N.K. (1981), "Inference from stratified samples: properties of the linearization, jacknife and balanced repeated replication methods" The Annals of Statistics, Vol 9, N.5, pp. 1010-1019.

Kitagawa E. M. (1955), "Components of a difference between two rates", Journal of the American Statistical Association, Vol. 50, pp. 1168-1194.

Kish L. (1965), "Survey Sampling", John Wiley & Sons, Inc.

Knottnerus P. (2012), "On aligned composite estimates from overlapping samples for growth rates and totals" Discussion paper CBS, Statistics Netherlands.

Knottnerus P. & Van Delden A. (2012), "On variances of changes estimated from rotating panels and dynamic strata", Survey Methodology, Vol 38, N° 1, pp. 43-52.

Knottnerus P. & Van Delden A. (2006), “Estimation of changes in repeated surveys and their significance”, https://www.iser.essex.ac.uk/ulsc/mols2006/programma/data/paper/Knottnerus.doc.

Kovar J.G., Rao J.N.K., Wu C.F.J. (1988), “Bootstrap and other methods to measures errors in survey estimates”, Statistical Society of Canada, pp. 25-45.

Laniel N. (1988), “Variances for a rotating sample from a changing population”, Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp. 246-250.

Nordberg L. (2000), “On variance Estimation for measures of change when samples are coordinated by the use of permanent random numbers”, Journal of Official Statistics, Vol. 16 No. 4,2000, pp. 363-378.

Moretti D., Pauselli C., Rinaldelli C. (2005), “La stima della varianza campionaria di indicatori complessi di povertà e disuguaglianza”, Statistica Applicata, Volume 17, n.4 pp 529-549.

Osier G., Berger Y., Goedemé T. (2013), “Standard error estimation for the EU–SILC indicators of poverty and social exclusion”, Eurostat Methodologies and Working papersce, Brussels, 5-7 March.

Osier G.& Perray P. (2016), “Variance estimators of annual levels and net changes for a defined set of LFS-based indicators”, Sogeti, Deliverable 1 – Methodological report.

Osier G. & Raymond V. (2017), “Development of methodology for the estimate of variance of annual net changes for LFS-based indicators”, Sogeti, Deliverable 1 - Short document with derivation of the methodology (FINAL).

Osier G. (2009), “Variance estimation for complex indicators of poverty and inequality using linearization techniques”, Survey Research Methods, Vol. No.3, pp. 167-195.

Wolter K. M. (1985), “Introduction to variance estimation”, Springer-Verlag. New York.

Wood J. (2008), “On the covariance between related Horvitz-Thompson estimators”, Journal of official statistics, Vol. 24. No. 1, 2008, pp.53-78.

Woodruf R. (1971), “A simple method for approximating the variance of a complicated estimate”, Journal of the American Statistical Association, 66(334), pp. 411-414.

Qualité L. & Tille Y. (2008), “Variance estimation of changes in repeated surveys and its application to the Swiss survey of value added”, Survey Methodology 34, issue 2, pp. 173-181.

Qualité L. (2008), “A comparison of conditional Poisson sampling versus unequal probability sampling with replacement”, Journal od Statistical Planning and Inference, pp. 1428-1432.

Quatember A. (2014), “The Finite Population Bootstrap - From the Maximum Likelihood to the Horvitz-Thompson Approach”, Austrian Journal of Statistics, Vol. 43, pp. 93-102.

Rao J.N.K. & Wu C.F.J. (1984), “Bootstrap Inference for sample survey”, Proc. Section on Survey Research Methods. American Statistical Association, pp. 106-112.

Rao J.N.K & Wu C.F.J. (1988), “Resampling Inference with complex survey data”, Journal of the American Statistical Association, pp. 231-241.

Righi P., Solari F., Falorsi S. (2005), “Stime ed errori. Note Metodologiche” Istat.

Särndal C.E., Swensson B., Wretman J. (1989), “The weighted residual technique for estimating the variance of the general regression estimator of the finite population total”, Biometrika, Vol. 76, n. 3, pp. 527-537.

Statistics Sweden (2003), “Samu: the system for co-ordination of frame populations and sample from the Business Register at Statistic Sweden”, Background Facts on Economic Statistics.

Sukhatme P.V. & Sukhatme B.V. (1970), "Sampling Theory of Surveys with Applications", Iowa State University Press, Ames, Iowa, USA. p.p. 27-29.

Tam S.M. (1984), “On covariance from Overlapping Sample”, The American Statistician, Vol. 38 (4), pp.288-289.

Verma V., Betti G. (2005), “Sampling errors and design effects”, Working Papers, n. 53, Dipartimento di Metodi Quantitativi, Università di Siena.

Valliant R. (1991), “Variance Estimation for Price Indexes from a Two-Stage Sample with Rotating Panels”, Journal of Business & Economic Statistics, Vol. 9, n. 4 (Oct.), pp. 409-422.

Zannella F. (1989), “Manuale di tecniche di indagine. Tecniche di stima della varianza campionaria”, Istituto Nazionale di Statistica.

Zardetto D. (2015), “ReGenesees: an Advanced R System for Calibration, Estimation and Sampling Error Assessment in Complex Sample Surveys”, Journal of Official Statistics”, Vol. 31, n. 2, pp. 177–203.