RESEARCH IN OFFICIAL STATISTICS

2002 EDITION

COPYRIGHT

Luxembourg: Office for Official Publications of the European Communities, 2002

ISBN 92-894-4432-0
ISSN 1725-1567
Cat. No. KS-CR-02-001-EN-N
Price (excluding VAT) in Luxembourg: EUR 17

Monographs of official statistics

Variance estimation methods in the European Union

1

EUROPEAN COMMISSION

THEME 1
General statistics

Europe Direct is a service to help you find answers to your questions about the European Union

New freephone number:

00 800 6 7 8 9 10 11

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (http://europa.eu.int).

Luxembourg: Office for Official Publications of the European Communities, 2002

ISBN 92-894-4432-0
ISSN 1725-1567

ACKNOWLEDGEMENTS

The project on variance estimation methods in the EU was managed and coordinated by Mr Håkan Linden and Mr Jean-Marc Museux of Eurostat Unit A4 'Research and development, methods and data analysis'.

This report has been produced by Ms Zoi Tsourti and Mr Giorgos Kouvatseas under the direction of the Eurostat Task Force on 'variance estimation'.

The European Commission gratefully acknowledges the valuable contributions of all participants.

Members of the Task Force:
Institut National de Statistique (Belgium); Statistics Finland (Finland); INSEE (France); Statistisches Bundesamt (Germany); Central Statistical Office (Ireland); ISTAT (Italy); Statistics Norway (Norway); Central Statistical Office (Poland); Statistics Romania (Romania); Statistics Sweden (Sweden); Swiss Federal Statistical Office (Switzerland); Office for National Statistics (United Kingdom).

Publication information:
This publication is available in English and can be ordered free of charge through the Eurostat Data Shops and the Office for Official Publications of the European Communities.

FOREWORD

Variance estimation has become a priority as more and more Commission Regulations require that the quality of the statistics be assessed. Sampling variance is one of the key indicators of quality in sample surveys and estimation. Sampling variance helps the user to draw better conclusions about the statistics produced, and it is also important information for the design and estimation phases of surveys.

However, due to the complexity of the methods used for the design and the analysis of the survey, like the sampling design, weighting and the type of estimators involved, the calculations are not straightforward. The literature on variance estimation is rich; however, no clear guidelines exist. This is mainly because all the methods compete, due to the existence of different simplifications or approximations.

Because of the necessity to offer solutions to the methodological problems encountered in the very specific field of variance estimation among the members of the European statistical system (ESS) a Task Force was set up by the Eurostat Working Group on the Assessment of Quality. The Task Force, composed of specialists from European national institutes, met four times and discussed solutions to many of the methodological problems encountered for sample surveys in the ESS. The meeting documents and the final report of the Task Force are available on the CIRCA interest group 'Quality in Statistics' (http://forum.europa.eu.int/Public/irc/dsis/Home/main).

This report has been produced in order to provide a large visibility to the work of the Task Force. It provides a summary of the currently available variance estimation methods, and general recommendations and guidelines endorsed by the Working Group on the Assessment of Quality for the estimation of variance for the common sampling procedures used at the European level. Not all the issues raised by the Task Force are tackled in this report. Some of them are being studied in research projects on variance estimation issues under the fifth framework programme of the European Commission, and results are not yet available.

The report aims to provide a framework for survey statisticians and methodologists when choosing an appropriate method for estimating sampling variability of their estimates. But it aims as well to address professionals when analysing survey data. In order to retain its value as a source of information on 'currently used variance estimation methods', this report has to be regularly updated.

Jean-Louis Mercy Head of Unit	Pedro Diaz Muñoz Director

CONTENTS

INDEX OF TABLES

1. INTRODUCTION

This report examines the issue of variance estimation of simple statistics under several sampling designs and estimation procedures. It especially focuses on two representative examples of household and business surveys, labour force survey (LFS) and structural business statistics (SBS) respectively. It has been produced in the frame of the project 'Estimation techniques statistics' which is Lot 4 of **2000/S 135-088090** invitation to tender. The main objective of the present work is to provide:

- a depository of the currently available variance estimation methods;
- general recommendations and guidelines for the estimation of variance with respect to the common sampling procedures (incorporating sampling design, weighting procedures as well as imputation) deployed at European level.

The compilation of the content of the reporting, its structuring and presentation has been performed, having in focus the aforementioned objectives as well as to address more effectively professionals involved in analysis of survey data.

The report has the following structure: in the second chapter a number of factors that affect variance and the procedure of its estimation (such as sampling design, weighting etc.) are presented and their effect commented. The several alternative variance estimation techniques are discussed in brief in chapter 3. A theoretical comparison of those can be found in that chapter. The several software packages that have been developed for variance estimation, in recent years are described in chapter 4. Finally, in chapter 5, some practical guidelines for the implementation of variance estimation under conditions common to surveys conducted in Europe (with respect to sampling design and weighting) are provided. The discussion on imputation is performed via a case study, since imputation, being survey- and variable-sensitive, requires ad-hoc treatment in each case. In order to illustrate the various issues that may arise in the context of variance estimation as well as the multiple ways for handling them, several topics of specific interest have, also, been included in this chapter. The issue of calculation of coefficients of variation is also addressed there. The notation used throughout the report is described in the annex.

1.1 Importance of variance estimation

The primary concern in all sample surveys is the derivation of point estimates for the parameters of main interest. However, equally important is the derivation of the variances of the above estimates. The importance of variance estimators, and corresponding standard errors, mainly lays on the fact that the estimated variance of any estimator is a main component of the quality of any estimator.

In brief, as noted in Gagnon et al. (1997), variance estimation:

- provides a measure of the quality of estimates;
- is used in the computation of confidence intervals;
- helps draw accurate conclusions;
- allows statistical agencies to give users indications of data quality.

The sampling variance is, indeed, one of the key indicators of quality in sample surveys and estimation. It indicates the variability introduced by choosing a sample instead of enumerating the whole population, assuming that the information collected in the survey is otherwise exactly correct. For any given survey, an estimator of this sampling variance can be evaluated and used to indicate the accuracy of the estimates.

Thus, indeed, variance estimation is a crucial issue in the assessment of the survey results. However, due to the complexity of the methods used for the design and the analysis of the survey, like the sampling design, weighting, the type of estimators involved etc. the respective calculations are not straightforward. The literature on variance estimation is rich, however no clear guidelines exist. This is mainly because all the methods compete, due to the existence of simplifications or approximations. The choice depends on experience, resources and institutional mentality. In this report some rough recommendations are provided.

2. DECISIVE FACTORS FOR VARIANCE ESTIMATION

The choice of an appropriate variance estimator depends on:

- type of sampling design (i.e. stratified, multi-stage, clustered etc.);
- type of estimator (i.e. weighting);
- type of non-response corrections (i.e. re-weighting, imputation);
- measurement errors;
- form of the statistics (linear: totals, means (for known population size); non-linear: means (for unknown population size), ratios…).

The degree of complexity of the aforementioned dimensions dictates the complexity of any survey, with the first two factors being the main contributors. As depicted in Wolter (1985) (see figure below), a sample survey can be regarded as 'complex' when a complex sample design is deployed (irrelevant of the type of estimator used) or, even in cases of simple designs accompanied by non-linear estimators.

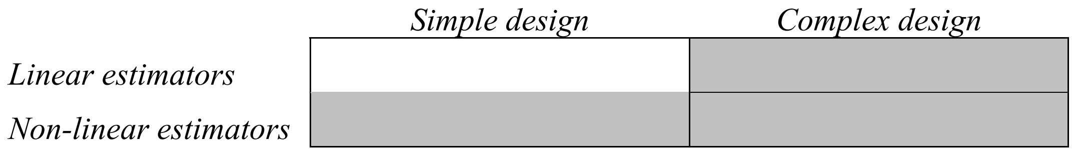

Figure 1: Graphical depiction of 'complex sample surveys'

The studies where there is a need for complex sample designs, those include namely varying probabilities and non-independent selections are very common. For example, in household surveys persons may be sampled from geographically clustered households. In this scheme, persons within the same geographical cluster have a higher probability of being sampled together than do persons in different clusters. Similarly, in some business surveys using local business units as sampling units, the samples are not independent because several local units could be sampled from the same enterprise and each enterprise has its own practices and procedures. Another type of dependence occurs when data are collected at several points in time. In addition to the complexities due to clustering, the probability of selecting a particular unit may vary depending on factors such as the size or location of the unit. For example, in a sample of businesses the probability of selecting a business may be proportional to the number of employees (πps or pps) (Särndal et al., 1992). These types of design features make analysis of the data more difficult.

When complex surveys are used to collect data, special techniques are needed to obtain meaningful and accurate analyses, since **ignoring the sample design and the adjustment procedures imposed on data leads to biased and misleading estimates of the standard errors**. Ignoring features such as clustering and unequal probabilities of selection leads to underestimation of variances, while disregarding of stratification usually leads to overestimation of variances.

Additional factors such as ease of computation, bias and sampling variability of the variance estimators, information required for calculation (and availability of such information on confidentially protected files) are important when one has to choose among more than one valid appropriate estimators.

In the sequel the impact of the main factors to the variance estimation are further elaborated.

2.1 Sampling design

The underlying sampling design of a sample survey is one of the most important factors that influence the size as well as the procedure required for the estimation of variances. More precisely, there are several components of sample designs that are related to the variance estimation:

— The number of stages of the sampling

Each additional stage of sampling adds variability to the finally derived estimates.

— The use or not of stratification of sampling units

Stratification is commonly used in practice in order to give a ‘more representative’ sample with respect to characteristics judged to be influential. This strategy, generally speaking, leads to a reduction in the total variance. In a stratified sample, the total variance is the weighted variance of each stratum.

— The use or not of clustering of sampling units

Clustering is, also, a usual strategy that aims at the reduction of the cost of a sampling survey. However, contrary to stratification, it, generally, leads to an increase of the total variance.

— The exact sample selection scheme(s) deployed

Finally, the sampling schemes that are deployed at each stage of a sampling design (equal or unequal selection probabilities), stratified or not, have a serious impact on the variance of any estimator and the way that it is estimated. Ignoring unequal selection probabilities of sampling units tends to an underestimation of standard errors.

The impact of these features to the process of variance estimation, of rather simple statistics such as totals (linear) as well as non-linear (e.g. ratios), is further discussed below in the current chapter.

2.1.1 Number of stages

In one-stage sample designs the situation is quite straightforward and the procedure of variance estimation depends only on the specific sampling scheme deployed as well as to whether stratification and/or clustering is used.

In the case of sampling designs with more than one stage the situation gets complicated due to the fact that there are more than one source of variation. **In each stage, the sampling of units (primary, secondary, and so on, up to ultimate) induces an additional component of variability**. In some cases (where all the other components of sampling and estimation are rather simple) a closed-form formula may be obtained, calculating the variance at each stage. However, the common practice is to approximately assess the variance by estimating the variability among primary sampling units, since this is the dominating component of total variance.

For example, in two-stage sampling we have two sources of variation: variation induced by the selection of primary sampling units (PSU) as well as variation resulting from the selection of secondary sampling units (SSU). Fortunately, the hierarchical structure of two-stage (or, accordingly, multi-stage) sampling designs leads to rather straightforward formulae for estimators and corresponding standard errors. Generally speaking, the variance estimation of a statistic can be decomposed into two parts, one part of PSU variance and another part of SSU variance (Särndal, et al. 1992). The exact form of the two variance-components depends on the sampling schemes utilized at each stage of the sampling.

2.1.2 Stratification

In stratified sampling the population is subdivided into non-overlapping subpopulations, called strata. Strata commonly define homogenous subpopulations, leading, thus, to a **reduction in the total variance**. From each stratum, a probability sample is drawn, independently from the other strata. The sampling design within each stratum could be the same or different from other strata. This independence among samples in different strata implies that any estimator as well as its corresponding variance estimator is simply the sum of the corresponding estimators within each stratum.

So, the problem of finding the most appropriate variance estimator for a single-stage stratified sampling reduces to the problem of the most appropriate variance estimator for the sampling designs deployed in each stratum.

2.1.3 Clustering

Clustering is a commonly used strategy in order to reduce the cost of a survey, with respect both to time and money. In a clustered sample design, the sampling unit consists of a group (cluster) of smaller units (elements). Sampling may be done for the clusters or the secondary sampling units. **In most cases, there is some degree of correlation (homogeneity) among elements within the same cluster, leading thus to an increase of the variance of any statistic (compared to the case of simple random sampling)**.

In clustered samples, the variance consists of two components: variance within clusters (which depends on the intra-correlation of elements) and variance among clusters. The total variation is dependent on the intra-correlation of elements and the variance of the elements if simple random sampling was deployed (for more details one may refer to Cochran, 1977 and Särndal et al., 1992).

Often, the issue of clustering is ignored in practice, leading to an underestimation of the variance. There are two approaches in order to account for clustering effect. First of all, if the sampling scheme and the type of estimator are simple, one could proceed into the analytic estimation of the two components of variance (within and between clusters) based on the calculation of the intra-correlation coefficient. However, this task is not straightforward in real situations. In such cases one could resort to appropriately adjusted resampling techniques, which will be further described in a later point of the report.

2.1.4 Sample selection schemes

- Simple random sampling

The simple random sampling is, as one may expect, the simplest case, leading to straightforward, closed-form exact formulae for the calculation of variances of estimators of linear forms. In case of non-linear statistics (such as ratios) linearisation is needed in order to derive closed-form, though approximate, formulae for the variance. This is not a hard task for statistics as simple as the ratio.

- Systematic sampling

Systematic sampling is a convenient sampling design, mainly used for the effort reduction in sample drawing that it offers. Moreover, whenever properly applied, it can incorporate any obvious or hidden stratification of the population, leading thus to greater precision than simple random sampling.

Unfortunately, one of the costs paid for the simplicity of the systematic sampling is that there is no unbiased estimator for the variance of statistics as simple as a total. So we have to resort to some biased estimators. Several such suggestions can be found in the literature. A comprehensive study is presented in Wolter (1985). The adequacy of these alternative estimators depends on the nature of the underlying population as well as to the logic followed during the compilation of the list (sorting) of the elements of the population.

The most common approach, used in practice, is to ignore the effect of systematic sampling and apply the formulae that hold for the case of simple random sampling. Another approach is based on the pseudo-stratification of the sample (that is, considering the systematic sample as a stratified random sample with 2 units from each successive stratum), while the generic notion of variance estimation via replication techniques, such as bootstrap, could be used here as well. In practice it is claimed that the error introduced from assuming simple random sampling is not significant, not justifying, thus, the additional burden imposed by the use of pseudo-stratification or replication methods. Moreover, according to Särndal et al. (1992), in multi-stage samples where systematic sampling is deployed in the final stage, the bias is not as serious as one might expect.

We make special reference to stratified systematic sampling, which is a form of probabilistic sampling, combining elements of simple random, stratified random, and systematic sampling, in an effort to reduce sampling bias (Berry and Baker, 1968). The impact of stratified systematic sampling to the variance within each stratum is analogous to the impact of simple systematic sampling to the variance, as previously discussed. This method will be more precise than stratified random sampling if systematic sampling within strata is more precise than simple random sampling within strata. As previously mentioned, in the case of systematic sampling a compromise has to be made in the variance estimation procedure. Since no unbiased variance estimation exists for this design, the simplifying assumption of simple random (or stratified, in this case) sampling may be adopted as long as the ordering of the sampling units before the systematic selection has been performed in such a way so as to lead to heterogeneous samples (as is usually the case). This restriction is imposed in order to prevent an underestimation of the variance. However, a more close approximation of the underlying sampling design can be achieved under the conceptual construction of a stratified two-stage clustered sampling. In this case the variance of a total can be estimated via the Jackknife linearisation method (Holmes and Skinner, 2000). This variance estimation method can also incorporate any weighting adjustments performed.

- Probability proportional-to-size sampling

Generally speaking, probability proportional-to-size (πpS or ppS) sample designs are rather complicated with complex second-order inclusion probabilities, leading, thus, to sophisticated formulae for variance estimation. Actually in most cases only approximations can be derived in practice. These are derived from corresponding simplifications in the sampling schemes (Särndal et al., 1992), so that one does not need to estimate second order probabilities. Of course, these approximations induce a component of bias in variance estimation.

For example, as illustrated in Wolter (1985), under the assumption of simple random sampling without replacement, the bias induced to variance estimation of a total (of πpS sampling) is given by the formula:

$$Bias\left(V(\hat{t})\right)=\hat{t}^2-\frac{2n}{n-1}\sum_{i=1}^{N}\sum_{j>i}^{N}\pi_{ji}\left(\frac{Y_i}{\pi_i}\right)\left(\frac{Y_j}{\pi_j}\right),$$

where π_i, π_j are the first-order inclusion probabilities of sample units i and j respectively, while π_{ij} is the second-order inclusion probability.

2.2 Calibration

Weighting is a common statistical procedure used in most survey data in order to improve the quality of the estimators in terms of both precision and non-response bias. Weighting can be used to adjust for the particular sample design, for (unit) non-response while it can also use additional auxiliary information in order to enhance the accuracy of estimates.

In business surveys highly skewed variables may damage the accuracy of the produced statistics. The effective use of auxiliary information, combining data from business surveys and administrative record systems, produces statistics of satisfactory quality. The use of ratio, combined ratio or regression estimators improve the accuracy of business statistics (Särndal et al., 1992, Deville and Särndal, 1992). The classical framework for the above estimators assumes that the auxiliary population information is correct and that the sample source is obtained using a known probability-sampling scheme. Moreover, ratio estimators serve extensively to adjust the produced statistics reducing the biases of non-coverage and non-response.

Note, however, that the above three estimators are biased but this bias is usually negligible (Kish et al., 1962). Moreover, the use of these calibration estimators may cause difficulties in the statistics production as the business surveys are multipurpose and multivariate and as a result, the model – based estimators may be suitable for some statistics but not for others.

A generic outline of a 'complete' weighting process is described below.

Stage 1: The major purpose of weighting is to adjust for differential probabilities of selection used in the sampling process. When units in the sample are selected with unequal probabilities of selection, expanding the sample results by the reciprocal of the probability of selection, i.e., sampling weight, can produce unbiased estimates. This type of weighting is, essentially, incorporated in the treatment of sampling design discussed in the previous section.

Stage 2: When a sample of subunits is used to get in touch with the sampling unit to which the sampled sub unit is associated (e.g. the members of a household or the local units of an enterprise), each subunit weight must be adjusted according to the number of subunits in the frame for that sampling unit. This is of importance for variables that are associated with the whole sampling unit.

Stage 3: Properly weighted estimates based on data obtained from a survey, would be asymptotically unbiased if every sampled unit agreed to participate in the survey, and every person responded to all items (questions) of the survey questionnaire. However, some non-response occurs even with the best strategies; thus adjustments are always necessary to minimize potential non-response bias. When there is unit non-response, non-response adjustments are made to the sample weights of respondents.

The purpose of the non-response adjustment is to reduce the bias arising from the fact that non-respondents are different from those who responded to the survey. The previously mentioned weights (of stages 1 and 2) are valid under the assumption that all units within a stratum respond (or not) with the same probability. However, when there is additional information available about the non-response more complicated and possibly better weights can be used.

Stage 4: If the frame contains auxiliary information about the sampling units, that is, variables that are correlated with at least some of the measurement variables of interest, this information could be used to improve the estimation. Calibration in its various forms has become an essential tool in surveys. Its justification is mainly based on the argument that calibration is a convenient way of improving the efficiency of estimates by exploiting external information. Other reasons for its use include the following (see also Mohadjer et al., 1996):

- balance (in the sense that after calibration, the sample 'looks like the population');
- target groups (if we need estimates for a specific subpopulation, we add the appropriate calibration variable);
- anchoring of estimates (a variable that is correlated with a calibration variable will tend not to 'drift' too far away, providing stability over time);
- convenience (e.g., regression-based calibration provides a simple way of getting a single weight per enterprise in a business survey, Lemaitre and Dufour (1987));
- consistency of estimates (in production systems, each sampled unit is given a unique final weight as part of the calibration process; as a result, estimates are consistent in the sense that, e.g., parts add up to totals);

- composite estimation (exploiting sample overlap over time to improve efficiency can be done via calibration).

All of the above weighting procedures can be easily incorporated in the generalised regression type of estimators (GREG).

An obvious advantage of the GREG estimator is the fact that it has been extensively studied in the literature. With respect to its variance estimator no analytic solution exists (Särndal et al., 1992). The simplest way is to approximate the variance using Taylor expansion. Unless the sampling design is complicated, Taylor linearisation is usually the preferred method.

At this point, the following clarification should be made. The finally derived weights do contain all the information necessary to construct point estimates. However, the weights alone do not give the extra information required to estimate variances, which are needed for inference. As an extreme example, for a design where each unit has equal probability of selection, the weights cannot reveal anything about the stratum memberships of the sample units; yet, with a stratified design, it is desirable to estimate the variance separately within each stratum. What additional information is required for variance estimation depends on the actual design of the survey, suitable approximations to that design as well as the weighting technique used.

2.3 Imputation

Non-response of sampled units is an inherent feature of all surveys. It can impair the quality of survey statistics by threatening the ability to draw valid inference from the sample to the target population of the survey. More particularly, non-response's impact on data and results is:

- increase of the sampling variance (since the finally derived sample is smaller than the originally planned sample size);
- introduction of bias to the estimates (in the case where the investigated behaviour of respondents deviates systematically from the corresponding behaviour of non-respondents).

There are two types of non-response: unit or item non-response. Each of these requires a different type of treatment. The knowledge of the type of 'amendments' performed is important for the proper evaluation of the variance estimators. Furthermore, another factor that determines the kind of variance estimator that is selected is the uniformity of the followed non-response approaches. If the level of uniformity is low then, usually, more computationally intense procedures are required. The most common method for dealing with unit non-response is that of weighting of the sample, which has been dealt with in the previous section.

Item non-response is most often dealt with by using imputation methods, that is 'prediction of the missing information'. The knowledge of the imputation methods used is essential not only for the final estimation of variance but even for the choice of the proper variance estimator, since imputation affects the variance of any estimated statistic. Along with the imputation method, flagging of the imputed values or, at least, the rate of imputation is the type of information that should always be provided with the data set.

A comprehensive review of the imputation techniques that are used at European level is provided by Patacchia (2000).

In the presence of imputed data, it is even more important to measure the accuracy of estimates, since imputation itself adds an additional process, which can be a source of errors. That is, the total error that we want to access consists of two components: the ordinary sampling error and the imputation error. Ignoring imputation and treating imputed values as if they were observed values may lead to valid point estimates (under uniform response within imputation classes) but it will unavoidably result in underestimation of the true variance (if the standard variance estimation methods are naively applied). In fact, as mentioned in Kovar and Whitridge (1995), even non-response as low as 5% can lead to an underestimation of the variance of order of 2–10%, while non-response rate of 30% may lead to 10–50% underestimation. This issue of underestimation of variance has also been illustrated in Full (1999).

The estimation of variance, which takes into account imputation, has the following benefits:

- gives better understanding of the impact of imputation;
- improves the estimation of total variance;
- permits better allocation of resources between a larger sample and improved verification and imputation procedures (according to the relative weights of variance due to sampling and variance due to imputation).

The choice of the most appropriate method for the estimation of variance in the presence of imputation depends on two primary factors:

- the underlying sampling design; and
- the imputation method employed.

Some other features that need to be taken into account are

- the imputation rate;
- whether more than one imputation method is used in our working data set; and
- non-negligible sampling fractions.

Generally speaking, the methods for variance estimation taking into account imputation, can be distinguished in the following categories:

- **Analytical methods**

Under this framework one can find the model-assisted approach of Särndal (1992) and Deville and Särndal (1991) and the two-phase approach of Rao (1990) and Rao and Sitter (1995).

- **Resampling methods**

The main replication techniques that have been developed for variance estimation are: jackknife technique, Rao (1991) and Rao and Shao (1992); Bootstrap, Shao and Sitter (1996); balanced repeated replication method (BRR), Shao et al. (1998). These methods are further discussed, in a more general context, in the next chapter.

- **Repeated imputations methods**

Some such methods are the ‘multiple imputation’ (Rubin, 1978, 1987) and the ‘all case imputation method’ (Montaquila and Jernigan, 1997).

An extensive review and comparison (both theoretically and empirically, via simulation) of the above methods is conducted in Rancourt (1998) and Lee et al. (1999). A comparative study with emphasis on multiple imputation is provided in Luzi and Seeber (2000).

Until recently no software existed for the calculation of variance taking into account the imputation deployed. Lately some attempts have been made to incorporate such a feature in some of the packages that are specially designed to deal with data from complex surveys and correctly calculating their variances.

Overall the choice of the variance estimation procedure that takes into account imputation, requires knowledge of the following information (Lee et al., 1999):

- indication of the response/non response status (flag);
- imputation method used and information on the auxiliary variables;
- information on the donor;
- imputation classes.

Not all the available methods for variance estimation require all the above information. However, flagging is necessary (Rancourt, 1996), otherwise only rough approximations are possible.

3. VARIANCE ESTIMATION METHODS

In standard textbooks for sampling, such as Cochran (1977), one may find straightforward exact analytic formulae for variance estimation of (usually simple) statistics under several sample designs. However, as sample designs get more and more complicated or deployed in more stages, no closed-form expressions exist for the calculation of variances. Even in cases of simple sample designs the utilization of advanced weighting procedures makes intractable the variance estimation formula of simple statistics, such as totals. In such cases, where no exact method exists for the calculation of unbiased estimates of the standard errors of the point estimates, the only alternative is to approximate the required quantities. Under the framework of analytic techniques, this is achieved by imposing simplifying assumptions (with respect to the sample design or the statistic to be variance-estimated). An alternative approach is based on replication methods. In the sequel we provide a short description of these methods in their general form in order to establish the theoretical framework that is related to the choice of the most suitable variance estimator for each circumstance.

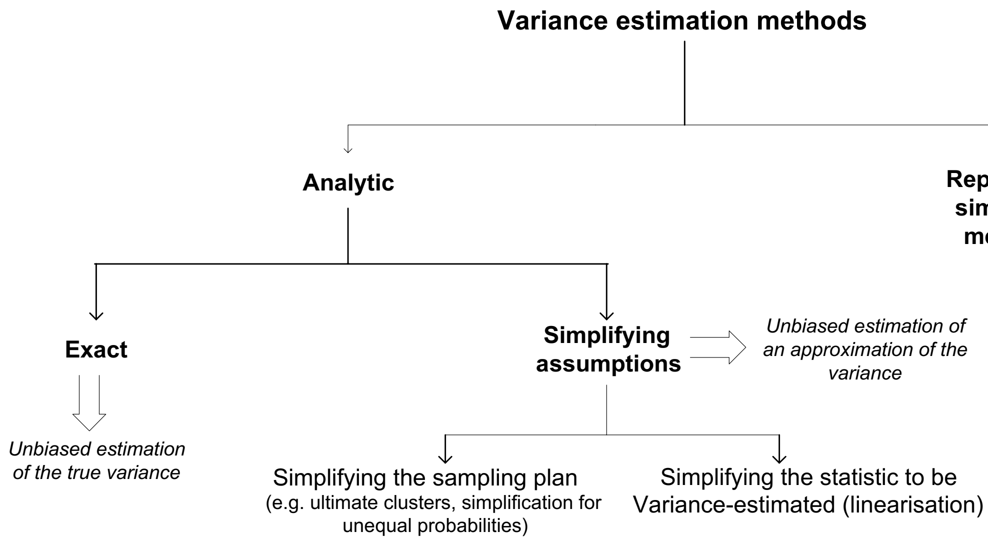

Figure 2: Overview of alternative variance estimation methods

3.1 Variance estimation under simplifying assumptions

3.1.1 Variance estimation under simplifying assumptions of sampling design

As previously mentioned, the complexities in variance estimation arise partly from the complicated sample design and the weighting procedure imposed. So a rough estimate for the variance of a statistic based on a complicated sample can be obtained either by ignoring the actual, complicated sample design used (unequal probabilities of selection, clustering, more than one stage of sampling); or proceeding to the estimation using the straightforward formulae of the simple random sampling or another similarly simple design.

However, generally speaking, the incorporation of sampling information is important for the proper assessment of the variance of a statistic. Since weighting and specific sample designs are particularly implemented for increasing the efficiency (and thus decreasing the variability) of a statistic, their incorporation in the variance estimation methodology is of major importance. For example, stratification tends to reduce the variability of a sample statistic, so if we ignore the design effect, the derived estimator will be upwards biased, overestimating the true variance of any statistic. Thus, the bias induced under this simplifying approach depends on the particular sampling design and should be investigated circumstantially. However, in general, this method is an indispensable one in common practice.

3.1.2 Variance estimation under simplifying assumptions of statistics (Taylor linearisation method)

The Taylor series approximation method relies on the simplicity associated with estimating the variance of a linear statistic, even with a complex sample design.

By applying the Taylor linearisation method, non-linear statistics are approximated by linear forms of the observations (by taking the first-order terms in an appropriate Taylor-series expansion)([1]). Extending the Taylor series expansion could develop second or even higher-order approximations. However, in practice, the first-order approximation usually yields satisfactory results, with the exception of highly skewed populations (Wolter, 1985)

Standard variance estimation techniques can then be applied to the linearised statistic. This implies that Taylor linearisation is not a 'per se' method for variance estimation, it simply provides approximate linear forms of the statistics of interest (e.g. a weighted total) and then other methods should be deployed for the estimation of variance itself (either analytic or approximate ones).

The Taylor linearisation method is a widely applied method, quite straightforward for any case where an estimator already exists for totals. However, the Taylor linearisation variance estimator is a biased estimator. Its bias stems from its tendency to underestimate the true value and it depends on the size of the sample and the complexity of the estimated statistic. Though, if the statistic is fairly simple, like the weighted sample mean, then the bias is negligible even for small samples, while it becomes nil for large samples (Särndal et al. 1992). On the other hand for a complex estimator like the variance, large samples are needed before the bias becomes small. In any case, however, it is a consistent estimator.

For more information on Taylor linearisation variance estimation method one may refer to Wolter (1985) and Särndal et al. (1992).

3.2 Variance estimation using replication methods

Replicate variance estimation is a robust and flexible approach that can reflect several complex sampling and estimation procedures used in practice. According to many

([1]) Note that Taylor series linearisation is, essentially, used in elementary cases, while influence function can be deployed in complex situations.

researchers, replication can be used with a wide range of sample designs, including multi-stage, stratified, and unequal probability samples. Replication variance estimates can reflect the effects of many types of estimation techniques, including among others non-response adjustment and post-stratification. Its main drawback is that it is computationally intensive. Especially in large-scale surveys, its cost in time may be prohibitively large. This will be made clearer in the sequel. Moreover, its theoretical validity holds only for linear statistics and asymptotics (which is a variation of linearisation).

The underlying concept of the replication approach is that based on the originally derived sample (full sample) we take a (usually large) number of smaller samples (sub-samples or replicate samples). From each sub-sample we estimate the statistic of interest and the variability of these 'replicate estimates' is used in order to derive the variance of the statistic of interest (of the full sample).

Let's denote by θ an arbitrary parameter of interest, $\hat{\theta} = f(data)$ the statistic of interest (the estimate of θ based on the full sample) and $v(\hat{\theta})$ the corresponding required variance. Then the replication approach assesses $v(\hat{\theta})$ by the formula

$$\hat{v}(\hat{\theta}) = c\sum_{k=1}^{G} h_k \left(\hat{\theta}_{(k)} - \hat{\theta}\right)^2$$

where

$\hat{\theta}_{(k)}$ is the estimate of θ based on the k-th replicated sample

G is the total number of replicates

c is a constant that depends on the replication method, and

h_k is a stratum specific constant (required only for certain sampling schemes)

There are several methods for drawing these 'replicate samples', leading thus to a large number of replication methods for variance estimation. The most commonly met in practice include the 'jackknife', 'bootstrap', 'balanced repeated replication' and 'random groups' along with their variants.

3.2.1 Jackknife estimator

The central idea of jackknife is dividing the sample into disjoint parts, dropping one part and recalculating the statistic of interest based on that incomplete sample. The 'dropped part' is re-entered in the sample and the process is repeated successively until all parts have been removed once from the original sample. These replicated statistics are used in order to calculate the corresponding variance. Disjoint parts mentioned above can be either single observations in a simple random sampling or clusters of units in multistage cluster sampling schemes. The choice of the way that sampling units are entered, re-entered in the sample (type and size of grouping) leads to a number of different expressions of jackknife variance. For example in JK1 method (which is more appropriate for unstratified designs) one sampling unit (element or cluster) is excluded each time, while in JK2 (more appropriate for stratified samples with two PSUs per stratum) and JKn (suitable for stratified samples with more than two PSUs per stratum) a single PSU is dropped from a single stratum in each replication.

It should also be noted that the jackknife method for variance estimation is more applicable in with-replacement designs, though it can also be used in without-replacement surveys when the sampling fraction is small (Wolter 1985). However, this is rarely the case when we are dealing with business surveys. The impact of its use in surveys with relatively large sampling fraction is illustrated, via simulation in Smith et al. (1998a), while, as mentioned in Shao and Tu (1995) the application of jackknife requires a modification — to account for the sampling fractions — only when the first stage sampling is without replacement. In any case, due to their nature, jackknife variance estimation methods seem to be more appropriate for (single or multistage) cluster designs, where in each replicate a single cluster is left out of the estimation procedure (neglecting, though, the finite population correction).

If the number of disjoint parts (e.g. clusters) is large, the calculation of replicate estimates is time consuming, making the whole process rather time-demanding in the case of large-scale surveys (Yung and Rao, 2000). So alternative jackknife techniques have been developed.

Jackknife linearised variance estimation is a modification of the standard jackknife estimator based on its linearisation. Its essence is that repeated recalculations of a statistic (practically numerical differentiation) are replaced by analytic differentiation. The result is a formula that it is easy to calculate. For example for stratified cluster sample the bias adjusted variance formula, presupposing sampling with replacement, is (Canty and Davison, 1999):

$$\hat{v} = \sum_{h=1}^{H} (1 - f_h) \cdot \frac{1}{n_h \cdot (n_h - 1)} \sum_{j=1}^{n_h} l_{hj}^2$$

The factor l_{hj} is the 'empirical influence value' for the jth cluster in stratum h. The calculation of l_{hj} is outlined in the appendix of Canty and Davison (1999). The effort required for calculating l_{hj} is based on the complexity of the statistic:

For the linear estimator in stratified cluster sampling:

$$\hat{\theta} = \sum_{h,j} y'_{hj}$$

where

$$y'_{hj} = \sum_{k} \omega_{hjk} \cdot y_{hjk}$$

is the sum of ys in every cluster j in each stratum h, and ω_{hjk} is the design weights

then

$$l_{hj} = n_h \cdot y'_{hj} - \sum_{j} y'_{hj}$$

For the ratio of two calibrated estimators, l_{hj} is:

$$l_{hj} = \frac{l_{hj}^{y} - \hat{\theta} \cdot l_{hj}^{z}}{1^{T} Wz}$$

where:

$$\hat{\theta} = \frac{1^{T} Wy}{1^{T} Wz}$$

while y and z are the vectors of the observations in the dataset and l_{hj}^{y}, l_{hj}^{z} and W are calculated from the data analytically.

Overall we can say that its main advantage is that it is less computationally demanding, while it generally retains the good properties of the original jackknife method. However, in case of non-linear statistics, it requires the derivation of separate formulae, as is the case with all linearised estimators. Therefore, its usefulness for complex analyses of survey data or elaborate sample designs is somewhat limited. For more details one may refer to Canty and Davison (1998, 1999) and Rao (1997), while an insightful application is made by Holmes and Skinner (2000).

3.2.2 Bootstrap estimator

The bootstrap involves drawing a series of independent samples from the sampled observations, using the same sampling design as the one by which the initial sample was drawn from the population and calculating an estimate for each of the bootstrap samples. Its utility in complex sample surveys has been explored in some particular cases. However, since the bootstrap technique was not developed in the frame of sampling theory, there are still some issues that need to be investigated such as the issue of non-independence between observations in the case of sampling without replacement as well as other complexities. In order to ensure an unbiased result the variance of the bootstrap estimator is multiplied with a suitable constant.

In the case of stratified sample designs, resampling is carried out independently in each stratum. Its main drawback is that it is too time consuming.

3.2.3 Balanced repeated replication method

The balanced repeated replication method (BRR) (or balanced half samples, or pseudoreplication) has a very specific application in cluster designs where each cluster has exactly two final stage units or in cases with a large number of strata and with only two elements per stratum. The aim of this method is to select a set of samples from the family of 2^k samples, compute an estimate for each one and then use them for the variance estimator in a way that the selection satisfies the 'balance' property (for a brief description see Särndal et al., 1992).

In the cases where the clusters have variable number of units, the division of them into two groups is required and thus modifications have been developed. For example for the stratified designs one has to treat each stratum as if it were a cluster, and to use divisions of the elements into two groups. However, where there is an odd number of elements in the stratum the results are biased, and ways of reducing this bias (but not eliminating it) are described in Slootbeek (1998). Recent research (Rao and Shao 1996) shows that only by using repeated divisions ('repeatedly grouped balanced half samples') can an asymptotically correct estimator be obtained.

Therefore the use of BRR with business surveys is typically difficult, as stratification is regularly used and the manipulation of both data and software becomes very difficult.

According to Rao (1997) the main advantage of BRR method over the jackknife is that it leads to asymptotically valid inferences for both smooth and non-smooth functions. However, it is not easily applicable for arbitrary sample sizes n_h like the bootstrap and the jackknife.

3.2.4 Random groups method

The random group method consists of drawing a number of samples (sub-samples) from the population, estimating the parameter of interest for each sub-sample and assessing its variance based on the deviations of these statistics from the corresponding statistic derived from the union of all the sub-samples. This technique is fully described in Wolter (1985), while it is also explored in 'Variance calculation software: evaluation and improvement (Supcom 1998 project, Lot 16)'. As mentioned therein, this was one of the first methods developed in order to simplify variance estimation in complex sample surveys.

Random groups method can be distinguished into two main variations, based on whether the sub-samples are independent or not. In practice, the common situation is that survey sample is drawn at once and random groups technique is applied in the sequel by drawing, essentially, sub-samples of the original sample. In such cases, we, almost always, have to deal with dependent random groups.

In the case of independent random groups, this technique provides unbiased linear estimators, though small biases may occur in the estimation of non-linear statistics. In case of dependent random groups, a bias is introduced in the results, which, however, tends to be negligible for large-scale surveys with small sampling fraction. In such circumstances the uniformity of the underlying sampling design of each sub-sample is a prerequisite for safeguarding the acceptable statistical properties of the random groups variance estimator.

3.2.5 Properties of replication methods

A common component of all the replication methods is the derivation of a set of replicate weights. These weights are re-calculated for each of the replicates selected, so that each replicate appropriately represents the same population as the full sample. This may be considered as a disadvantage as additional computational power is required to carry out all these calculations. However, this requirement is balanced by the unified formula for calculating variance, as no statistic-related formula is needed, since the approximation is a function of the sample, not of the estimate.

The primary reason for considering the use of resampling methods, which justify the additional computational burden they impose, is their generic applicability and adaptability. Indeed, these methods can be used without major changes irrespectively of the sampling design used, the type of estimator whose variance one tries to estimate and adjustments imposed. More importantly this can be done rather easily (just by calculating appropriate weights) compared to the modifications that a standard analytic variance estimator needs in order to take into account such information.

Some other appealing features of replication methods are their simplicity (its main essence is easily understood even among data users without special training in variance estimation), their sound theoretical basis (i.e. they are justified in the context of design-based as well as model-based approach), their easy application to domain estimates as well as their ability to consistently deal with missing data. The replication approach can also play a role in safeguarding confidentiality.

A more thorough examination of replication methods can be found in Morganstein (1998) and Brick et al. (2000).

3.3 Comparison of the methods

The appropriateness of each of the aforementioned variance estimation methods depends on the sampling design and the adjustments that are deployed in each case. However some general comments may be derived for classes of sampling designs or weighting methods.

Undeniably, exact formulae constitute the 'best' approach, but they are not available, or they are too difficult to be derived, in many practical cases of complex surveys. As long as the use of simplifying assumptions (of sample design) is concerned, we could mention that their ad-hoc use is, generally speaking, rather unsafe in cases that we are not certain that any effect of sample design or weighting does not significantly affect the precision of estimates. Moreover, the bias depends on the particular sampling design and should be investigated circumstantially.

Replication methods, along with the Taylor linearisation one, have been compared both theoretically as well as empirically. Theoretical studies (Krewski and Rao, 1981, Rao et al., 1992) have shown that linearisation and replication approaches are asymptotically equivalent. Furthermore, simulation studies (Kish and Frankel, 1974, Kovar et al., 1988, Rao et al., 1992, Valliant, 1990) show that both methods, in general, lead to consistent variance estimators. In particular, jackknife methods (among the replication methods) have similar properties with the linearisation approach, while the properties of balanced repeated replications and bootstrap techniques (both of which belong to the replication approach) are comparable.

The equivalence (even if it is only asymptotically) of the two approaches implies that criteria other than the precision of the methods should be deployed in order to choose a method. So in the case of rather simple situations of sample designs and estimation features, linearisation may be simpler to interpret and less time demanding. However, in case of complex survey design and estimation strategies, replication methods are equivalently flexible.

As it is quoted in Wolter (1985), summarising findings from five different studies concludes: '… we feel that it may be warranted to conclude that the TS [Taylor series] method is good, perhaps best in some circumstances, in terms of the MSE and bias criteria, but the BHS [balanced half-samples] method in particular, and secondarily the RG [random groups] and J [jackknife] methods are preferable from the point of view of confidence interval coverage probabilities'.

Resampling-based variance estimates have been shown to be useful in certain specialised problems (see e.g. Canty and Davison, 1999). However, since it is generally unclear how to extend resampling methods beyond even stratified random sampling, they should be applied with extreme care and, in general, for the analysis of complex surveys. For example (Bell, 2000), jackknife variance estimators are usually justified in a context of a stratified sample and assuming probability proportional to size (pps) or simple random sampling of clusters within strata. For the group jackknife method this justification can be found in Kott (1998). In the stratified sampling setting with a fixed number of strata, bootstrap procedures are available that provide improvements over classical approaches for constructing confidence intervals based on the normal approximation. However, the improvements are of second order and are generally only noticeable when the sample sizes are small. Moreover, in the case where there are an increasing number of strata, replication methods are likely to loose their appealing features as they provide minor asymptotic improvement over the standard normal approximation.

4. SOFTWARE FOR VARIANCE ESTIMATION

Recently, there has been a rapid growth in software market for software appropriate for analysing data (and so properly estimating variances) under complex survey designs and taking into account the adjustments (weighting, imputation) applied in such kind of data. These software packages are either add-on modules in already existing statistical packages or they may be stand-alone statistical software. A comprehensive review and comparison of several software for variance estimation can be found in 'Model quality report in business statistics, Vol II' (Smith et al, 1998a,b).

In the sequel a collection of software allowing variance estimation for survey data are presented, while in the subsequent table their main features are summarise.

- **Bascula**

Bascula (currently in its version 4) is a software package, part of the Blaise System for computer-assisted survey processing, for weighting sample survey data and performing corresponding variance estimation. It supports incomplete post-stratification and GREG weighting, while the available sampling designs include stratified one or two-stage sampling (multi-stage stratified sampling can be also hosted if replacement is used in the first stage). Variance estimation can be performed for totals, means and ratios based on Taylor linearisation and/or balanced repeated replication (BRR).

- **Caljack**

Caljack is an SAS macro, developed in Statistics Canada, in the framework of specific surveys. It is an extension of the SAS macro Calmar in order to cover the need for variance estimation. It covers stratified sample surveys (of elements or clusters), but the design weights need to be computed beforehand and introduced ready to Caljack. It can proceed to variance estimation of statistics such as totals, ratios (subsequently, means and percentages) and differences of ratios based on the jackknife technique. It provides all the calibration methods that are available in Calmar, that is, the family of calibration weights.

- **CLAN**

CLAN, developed in Statistics Sweden, is a program of SAS-macro commands. Taking into account the sampling design (stratified or clustered), it provides point estimates and standards errors for totals as well as for means, proportions, ratios or any rational function of totals (for the whole population or domains). πps sampling can only be approximated, while the only two-stage sampling that can be used is the one with simple random sampling of SSU. Incorporation of auxiliary information in the estimation is supported via GREG-type estimators (which also include complete or incomplete post-stratification). With respect to the treatment of unit non-response it allows for specific non-response models (by defining homogeneity response groups) as well as incorporation of sub-sampling of non-respondents.

The standard errors are calculated using the Taylor linearisation method for variance estimation.

- **Clusters**

Clusters is a command-driven stand-alone program, operating in DOS environment. It facilitates sample designs in the framework of stratified multistage cluster sampling, addressed through the ultimate cluster-sampling model. It offers sampling error estimates for totals, means, proportions and ratios for the whole population as well as for separate domains. Apart from standard errors it can also automatically compute coefficients of variation and design effects. However, it doesn't allow the incorporation of weights other than the sampling ones.

Standard errors are calculated using the Taylor linearisation method for variance estimation.

Clusters was originally designed in the framework of World Fertility Survey and later updated by V. Verma and M. Price.

- **Generalised estimation system (GES)**

GES, developed in Statistics Canada (Estevao et al. 1995), is an SAS-based application, running under SAS, with a windows-type interface. It can take into account stratified random sampling designs, elements or clusters, but not multi-stage (with more than one stage) designs and provide, accordingly, variance estimators for totals, means, proportions or ratios (for the whole population or domains). Methods of variance estimation available include Taylor linearisation and jackknife techniques. These techniques, apart from the sampling design, can also incorporate information of auxiliary variables in the weighting procedure. That is, it accommodates, apart from H-T, GREG type estimators.

- **Generalized software for sampling errors (GSSE)**

GSSE is a generalized software in SAS environment, developed within ISTAT, mainly devoted to the calculation of statistics and corresponding standard errors of data from sample surveys (for the whole population or domains). This software can take into account the sampling features of stratification (with or without replacement), probability proportional to sizes sampling, clustering and multiple-stages. In the case of multi-stage sampling, as other software, estimated variance is based solely on PSU variance. Weights for non-response adjustment, complete or incomplete post-stratification can be incorporated via the GSSW companion software.

The standard errors are calculated using the Taylor linearisation method for variance estimation.

- **Imputation and variance estimation software (IVEware)**

Similarly to GES and GSSE, IVEware is a SAS-based application, running under SAS, with a windows-type interface. It accounts for stratified random sampling designs, elements or cluster, but not multi-stage (with more than one stage) designs.

Variance estimators can be obtained for means, proportions and linear combinations of these, using Taylor linearisation procedure as well as for the parameters of linear, logistic, Poisson and polytomous regression (using the jackknife technique). However, no special technique is available for the adjustment of weights. One may incorporate in the estimation previously calculated (probably in another software) weights. But this leads to valid variance estimations only in the case of complete post-stratification, where the post-strata coincide with the strata themselves. IVEware may additionally perform imputation, but cannot incorporate this feature into variance estimation.

IVEware has been developed under the Survey Methodology Program in Survey Research Program, Institute for Social Research, University of Michigan by Raghunathan, T.E., Solenberger, P.W., Van Hoewyk, J.

- **PC CARP**

PC CARP (Iowa State University) is a stand-alone package used to analyse survey data. Using Taylor linearisation method it can compute variances of totals, means, quantiles, ratios, difference of ratios, taking into account the sampling design (it supports multistage stratified samples). Its companion Postcarp can provide point estimates incorporating post-stratification weighting.

- **Poulpe**

The program Poulpe (Programme optimal et universel pour la livraison de la précision des enquêtes), based on SAS, has been developed by INSEE and it can incorporate sampling features such as stratification, clustering or multistage-sampling, while it can also approximate variances in case of ppS sampling. Poulpe cooperates with Calmar and it takes the GREG weights provided by the latter in order to estimate variance of totals, ratios etc. based on the Taylor linearisation technique.

Ref: 'Variance estimators in survey sampling' Goga, C. (ENSAI, France) (In CIRCA: Quality in Statistics\SUPCOM projects\POULPE\calculation_poulpe)

- **SAS procedures**

Apart from the SAS macro commands that have been developed to cover specific needs of NSIs, such as CLAN or GSSE, the latest versions of the SAS statistical package (from version 7 and onwards) make provision for the valid estimation of standard errors of simple descriptive statistics as well as linear regression models. This is accomplished via the 'Surveymeans' and 'Surveyreg' SAS procedures.

Surveymeans procedure estimates descriptive statistics and their corresponding standard errors taking into account the sampling design (stratification or clustering) and possible domain estimation, while Surveyreg performs linear regression analysis providing variance estimates for the regression coefficients as well as any linear function of the parameters of the model (taking into account the specified sampling design and weights). However, in the case of clustered or multi-stage sampling, the variances are estimated based only on the first-stage of the sampling leading to an underestimation of variance. Moreover, in general, variance estimation is based on the assumption of sampling with replacement, which usually is not the case in practice.

This may lead to an overestimation of variance, which is, however, deemed to be negligible, especially in surveys with small first-stage sampling fraction.

No special technique is available for the adjustment of weights. However, one may incorporate in the estimation previously calculated (probably in another software) weights. This leads to valid variance estimations only in the case of complete post-stratification, where the post-strata coincide with the strata themselves.

Standard errors are calculated using the Taylor linearisation method for variance estimation.

- **STATA**

STATA is a complete statistical software package and the survey commands are part of it. STATA can correctly (i.e. taking into account the sampling design, stratified, clustered or multi-stage) estimate the variance of measures such as totals, means, proportions, ratios (either for the whole population or for different subpopulations) using the Taylor linearisation method. There are also commands for jackknife and bootstrap variance estimation, although these are not specifically oriented to survey data. Other analyses (such as linear, logistic or probit regression) can be performed by taking into account the sampling design (in the estimation of corresponding variances). However, STATA does not allow for variance estimation properly adjusted for post-stratification. (One may use in the estimation previously calculated weights, which leads to valid variance estimations only in the case of complete post-stratification, where the post-strata coincide with the strata themselves.)

- **Sudaan**

Sudaan is a statistical software package for the analysis of data from sample surveys (simple or complex). Though it uses SAS-language and has similar interface, it is a stand-alone package. It can estimate the variance of simple quantities (such as totals, means, ratios in the whole population or within domains) as well as more sophisticated techniques (parameter estimates of linear, logistic and proportional hazard models). The available variance estimation techniques include the Taylor linearisation, jackknife and balanced repeated replication. Again, weighting adjustments are not generally supported.

- **WesVar**

WesVar is a package primarily aiming at the estimation of basic statistics (as well as specific models) and corresponding standard errors from complex sample surveys utilizing the method of replications (balanced repeated replication, jackknife and bootstrap). Domain estimation and analysis of multiply-imputed data sets are accommodated. It can incorporate sample designs including stratification, clustering and multi-stage sampling. Moreover, it can calculate (and take into account in the variance estimation) weights of non-response adjustments, complete or incomplete post-stratification.

In the following table, a comparative presentation of the aforementioned software for variance estimation with respect to the main features of them, are presented.

Table 1: Comparative presentation of variance estimation software

Software Features		*Bascula*	*Caljack*	*CLAN*	*SAS Procedures*	*Clusters*	*GES*	*GSSE*	*IVEware*	*PC CARP*	*POULPE*	*STATA*	*Sudaan*	*WesVar*
VE methods	*Taylor's Linearis.*	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	Jackknife		✓				✓		✓				✓	✓
	Bootstrap													✓
	BRR	✓											✓	✓
Type of parameters for which variance can be estimated	*Means*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Totals	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓
	Percentages	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Ratios	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓
	Built-in functions			✓				✓	✓					✓
	Regression				✓				✓	✓		✓	✓	✓
Sampling designs adopted	*Simple random*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Stratified	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	2-stage	✓		✓([2])	✓	✓		✓		✓	✓	✓	✓	✓
	>2 stage	✓([3])			✓	✓		✓		✓	✓	✓	✓	✓
	Clustered	~([4])	✓	✓	✓	✓	✓	✓	✓	~	✓	✓	✓	~

([2]) It can provide valid variance estimators only if: i) we are referring to the estimation of a total, and ii) SSU are selected with simple random sampling, iii) H-T estimators only.

([3]) Only in with-replacement designs.

([4]) Approximately.

Software Features		*Bascula*	*Caljack*	*CLAN*	*SAS Procedures*	*Clusters*	*GES*	*GSSE*	*IVEware*	*PC CARP*	*POULPE*	*STATA*	*Sudaan*	*WesVar*
	pps	~	~	~	~	~	~	✓	~	~	~	~	~	~
Adjustments	*Complete poststratification*	✓	✓	✓	✓(5)		✓	✓	✓(6)		✓	✓(7)	✓(8)	✓
	Incomplete post-stratification	✓	✓	✓			✓	✓			✓			✓
	GREG	✓	✓	✓			✓	✓						✓
	Treatment for unit non-response			✓										✓
Interface	*Commands*		✓	✓	✓	✓			✓	✓	✓	✓	✓	
	Menu	✓					✓	✓			✓			✓
Cost	*Freeware*		✓(9)	✓(10)		✓(11)		✓(12)	✓		✓(13)			
	Commercial	✓			✓		✓			✓		✓	✓	✓

(5) The weights have to be calculated beforehand. Variance estimates are valid only if post-strata coincide with the strata.

(6) The weights have to be calculated beforehand. Variance estimates are valid only if post-strata coincide with the strata.

(7) The weights have to be calculated beforehand. Variance estimates are valid only if post-strata coincide with the strata.

(8) The weights have to be calculated beforehand. Variance estimates are valid only if post-strata coincide with the strata.

(9) It is for internal use.

(10) It requires SAS.

(11) To organizations and individuals in developing countries.

(12) It is for internal use.

(13) It is for internal use.

5. SOME PRACTICAL GUIDELINES

In chapter 3, a number of different variance estimation methods have been described. Among them, there is not an 'optimal' method, since the choice of the most appropriate one, in each case, is dependent on a series of criteria. As previously mentioned, the most critical factors are the type of sampling design, estimator, weighting adjustment and imputation performed. Additional factors such as ease of computation, bias and sampling variability of the variance estimators, information required for calculation (and availability of such information on confidential files) are important when one has to choose among more than one valid appropriate estimators.

In the present chapter, an effort is made to organise the several variance estimation methods, with respect to the usual sampling designs that are used in practice for the surveys conducted in EU members. However, even with good criteria and with the use of advanced software, variance estimation is not an easy task and decisions based on experience have to be taken.

We do not distinguish with respect to stratification, since this is an issue that can be easily incorporated in any situation (the formulae mentioned below mainly refer to the stratified case, but unstratified formulae can be easily derived if one sets the number of strata equal to 1).

Moreover, the issue of incorporation of imputation into the estimation of variance is treated separately, since its treatment is rather complicated and is dependent on the exact imputation scheme deployed. For this reason, we illustrate the treatment of imputation into a specific common case.

In section 5.3, a number of specific issues related to the variance estimation procedure are discussed (outliers, domain estimation, one-unit per strata problem, field substitution). Finally, in section 5.4 some general guidelines are provided for the calculation of coefficients of variation (a prominent alternative for expressing variability) at national as well as EU level.

5.1 Some suggestions for variance estimation

5.1.1 One-stage designs

5.1.1.1 *Sampling of elements*

- Simple random sampling

This is the simplest case, where closed-form formulae for variance estimation can be easily derived. More particularly in case of **linear** statistics, such as totals, exact solutions can be obtained (when no calibration or, simply, complete post-stratification is performed, else approximate formulae are required), while for **non-linear** statistics (e.g. ratios) one can develop approximate closed-form formulae based on the Taylor linearisation technique.

In the sequel we provide indicatively some such formulae for the case of a total, mean, percentage and ratio estimator, distinguishing whether weighting for calibration is performed or not. The notation that is used is explained in more detail in the Appendix.

(i) No weighting is used (apart from the sampling weights)

— A total t is estimated by the formula:

$$\hat{t} = \sum_{h=1}^{H} \hat{t}_h = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi}$$

while its variance $V(\hat{t})$ can be exactly estimated as:

$$\hat{V}(\hat{t}) = \sum_{h=1}^{H} \hat{V}(\hat{t}_h) = \sum_{h=1}^{H} \frac{N_h^2}{n_h} \cdot \left(1 - \frac{n_h}{N_h}\right) \cdot s_h^2 \text{, where } s_h^2 = \sum_{i=1}^{n_h} \frac{1}{n_h - 1} (y_{hi} - \bar{y}_h)^2$$

— In a similar fashion, the mean μ of the corresponding total t is estimated as:

$$\hat{\mu} = \frac{1}{N} \sum_{h=1}^{H} \sum_{i=1}^{n_h} \frac{N_h}{n_h} \cdot y_{hi} = \frac{1}{N} \hat{t}$$

$$\hat{V}(\hat{\mu}) = \frac{1}{N^2} \hat{V}(\hat{t}) = \frac{1}{N^2} \sum_{h=1}^{H} \frac{N_h^2}{n_h} \cdot \left(1 - \frac{n_h}{N_h}\right) \cdot s_h^2$$

— Any percentage can be regarded as the mean of a binary variable and thus it can be estimated as:

$$\hat{p}_A = \hat{p} = \frac{1}{N} \sum_{h=1}^{H} \frac{N_h}{n_h} n_{A(h)} = \frac{1}{N} \sum_{h=1}^{H} N_h \cdot \hat{p}_{A(h)}$$

with sample variance

$$\hat{V}(\hat{p}) = \frac{1}{N^2} \sum_{h=1}^{H} \frac{N_h (N_h - n_h)}{(n_h - 1)} \cdot \hat{p}_{a(h)} (1 - \hat{p}_{a(h)})$$

— The estimation of a ratio (as well as its variance) is referred to the estimation of the totals that constitute the ratio. That is,

$$\hat{r} = \frac{\hat{t}^y}{\hat{t}^z} \text{ and } \hat{V}(\hat{r}) \cong \frac{1}{(\hat{t}^z)^2} \left[\hat{V}(\hat{t}^y) + \hat{r}^2 \cdot \hat{V}(\hat{t}^z) - 2 \cdot \hat{r} \cdot \hat{C}(\hat{t}^y, \hat{t}^z) \right]$$

where $\hat{C}(\hat{t}^y, \hat{t}^z)$ is the estimated covariance between the two total estimators and it is estimated as:

$$\hat{C}(\hat{t}^y, \hat{t}^z) = \sum_{h=1}^{H} \left\{ \frac{N_h^2}{n_h (n_h - 1)} \cdot \left(1 - \frac{n_h}{N_h}\right) \cdot \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)(z_{hi} - \bar{z}_h) \right\}$$

➢ Note:

Apart from developing formulae and calculating variances by his/her own, one could use almost all of the aforementioned software for the estimation of variance in this framework. Note, here that WesVar uses only replication methods and does not provide for Taylor linearisation approach. Replication approach, though provides also valid approximation for this simple case, it is not suggested due to its additional computational burden.

(ii) Further weighting is used

As it has been mentioned in a previous section, in many cases, apart from the sampling design, more information is available and thus can be incorporated in the estimation procedure. The most usual estimator in this general case is the generalised regression (GREG) estimator. A more systematic study of GREG estimator and its properties can be found in Särndal et al. (1992). This estimator is not unbiased. It is only approximately unbiased for large sample sizes, while Taylor linearisation is required to derive approximate formulae for variance estimators, since the analytic forms are too complicated.

— In the case of total:

$$\hat{t} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{i=1}^{n_h} g_{hi} \cdot y_{hi}$$

The standard linearisation variance estimator (Särndal et al., 1992) is

$$\hat{V}(\hat{t}) = \sum_{h=1}^{H} \frac{N_h^2}{n_h} \cdot \left(1 - \frac{n_h}{N_h}\right) \cdot \frac{1}{n_h - 1} \sum_{i=1}^{n_h} g_{hi}^2 \left(y_{hi} - x_{hi}^T \cdot \hat{B}\right)^2$$

This type of variance estimator can be implemented in the most popular (specialised) software such as STATA, Sudaan and SAS (using the CLAN module).

— One may notice that the estimation of the mean is essentially the estimation of the corresponding total divided by the number of units in the population. This relationship holds irrespectively of the type of total estimator used. That is, if we want to estimate the mean of a variable y using the 'generalized regression' approach, it suffices to calculate $\hat{t}$, since then

$$\hat{\mu} = \frac{1}{N} \hat{t}$$

Accordingly, the variance estimator would be

$$\hat{V}(\hat{\mu}) = \frac{1}{N^2} \hat{V}(\hat{t})$$

So, all the comments applicable to variance estimation of totals apply here as well.

- Note:

In case that the size of population of interest, *N*, is regarded as unknown and, thus, it needs to be estimated, one should refer to the case of estimation of a ratio.

— Any percentage is essentially the mean of a dichotomous (dummy) variable. So the GREG estimator of a percentage can be derived in analogous manner as the GREG estimator of a mean.

— The ratio estimator, as shown in the simple case, is the ratio of the two total estimators (which can be derived as described above), while the corresponding variance is a function of the total estimators, their variance and a covariance term. That is, the additional effort that is required here is the estimation of the covariance term as described in the approximate formula below:

$$\hat{r} = \frac{\hat{t}^y}{\hat{t}^z},\ \hat{V}(\hat{r}) \cong \frac{1}{\left(\hat{t}^z\right)^2}\left[\hat{V}\left(\hat{t}^y\right) + \hat{r}^2 \cdot \hat{V}\left(\hat{t}^z\right) - 2 \cdot \hat{r} \cdot \hat{C}\left(\hat{t}^y, \hat{t}^z\right)\right]$$

$$\hat{C}\left(\hat{t}_y, \hat{t}_z\right) = \sum_{h=1}^{H} \frac{N_h^2}{n_h} \cdot \left(1 - \frac{n_h}{N_h}\right) \cdot \frac{1}{n_h - 1} \sum_{i=1}^{n_h} g_{y(hi)} g_{z(hi)} \left(y_{hi} - x_{y(hi)}^T \cdot \hat{B}_y\right)\left(z_{hi} - x_{z(hi)}^T \cdot \hat{B}_z\right)$$

➢ Note:

The aforementioned approximate formulae for variance estimation are based on the Taylor linearisation methods. However, an alternative approach (suggested by Morganstein, 1998) is the use of replication methods. More particularly:

◊ for the case of simple random sampling, without stratification, jackknife technique (JK1, in particular) is suggested,

while

◊ for stratified random sampling BRR and jackknife appear to be the main alternatives:

— if 2 units are selected per stratum JK2 and BRR are suggested;

— if more than 2 units are selected per stratum (with a small number of strata), JKn technique could be used;

— if the sample consists of a large number of units, then the above suggested replication methods will generally tend to be too time consuming. In such case, one could proceed to aggregation of the units into wider 'groups of units'.

✍ **Examples**

— In Labour Force Survey (LFS) in Luxembourg, where simple random sampling is deployed, a closed form formula is used for variance estimation, that takes into account the effect of post-stratification.

— In Continuing Vocational Training Survey (CVTS2) conducted in European level the sampling method used as in most business surveys, is one-stage stratified random sampling of enterprises (or local units). For the estimation of variance the use of Taylor linearisation method, as implemented in CLAN software, has been suggested (Quantos, 2001).

- Systematic sampling

As mentioned in a previous section, in the case of systematic sampling there is no unbiased variance estimate. The strategies that could be followed here (whether for linear or non-linear statistics) are:

— The most common approach used is to ignore the issue of systematic sampling, and proceed as if simple random sampling has been used (refer to the above paragraph). Of course, in such a case, the analyst should be aware that a potential overestimation of the obtained variance exists.

Alternatively:

— In case that there is enough available information about the listing of sampling units, pseudo-stratification could be deployed.

— According to Holmes and Skinner (2000), the jackknife linearisation (performed, for example, in STATA) is a promising approach, more efficient than Taylor linearisation approximation.

— Replication methods could be used, incorporating the effect of systematic choice of sampling units. The guidelines, among the several alternative replication methods are identical to those mentioned above for the case of simple random sampling.

✍ **Examples**

— Statistics Finland, in the framework of LFS, utilises a systematic, unstratified, sample design. For the estimation of variance of statistics of main interest, the assumption of simple random sampling is adopted and Taylor linearisation is used within CLAN software.

(Ref: 'Questionnaire from Statistics Finland', presented in 3rd Meeting of TF-VE)

— In the LFS of Sweden, under stratified systematic sampling, Taylor linearisation is also used, ignoring the feature of systematic drawing of units.

— In the GFSO (NSI of Germany), variance estimation of trade statistics (derived from an one-stage stratified systematic sampling of elements) is based on the assumption if stratified random sampling and the corresponding closed-form formulae (of Taylor linearisation approach)

(Ref: Practice of variance estimation at the German Federal Statistical Office, Presented in the 2nd Meeting of TF-VE, Doc. Eurostat/A4/Quality/Variance estimation/01/GFSO).

- 'ppS' sampling

Generally speaking, probability proportional-to-size (πpS or ppS) sample designs are rather complicated with complex second-order inclusion probabilities, leading, thus, to sophisticated formulae for variance estimation. Actually in most cases only approximations can be derived in practice. These are derived from corresponding simplifications in the sampling schemes (Särndal et al., 1992).

The suggested courses of action are:

— If the other features of a survey comply with the features that are provided in GSSE, one could use this software for variance estimation.

— An appropriate replication method could be developed, incorporating all the characteristics of the sampling procedure.

— An alternative approximate solution is the use of jackknife linearisation solution, which is currently used by Portugal in the framework of LFS (though in a two-stage design, which will be further discussed in the sequel).

— Further simplifications to the sampling design could be used, introducing though bias to the estimates.

5.1.1.2 *Sampling of clusters*

As mentioned in section 2.1.3, the clustering of sampling units, by inducing correlation among them, tends to an increase in variance.

For the case of simple random sampling of clusters (and not elements, now) closed form formulae for simple linear statistics can be obtained. However, since the situation is rather more complicated than element sampling, it is suggested (for simple random as well as other schemes) that one uses one of the software that is able to incorporate clustering([14]). Of course the choice of the software must take into account and the specific sampling scheme that is used for the selection of clusters, as has been discussed in 5.1.1.1, as well as on the decision of the most appropriate variance estimation technique.

As far as the particular method used is concerned, the choice is made between linearisation and replication methods.

— Taylor linearisation method could be used.

— For replication methods, we should mention that bootstrap and random groups can, generally, be adapted to any sampling design, though this task may become too cumbersome and time-consuming sometimes. So, for clustered sample designs in particular, jackknife (as well as its linearisation form) is more appropriate technique. Morganstein (1998) provides a detailed description of how these methods can be implemented in practice in WesVar software.

✍ **Examples**

— In the ONS (Office of National Statistics, of UK), the labour force survey is implemented using one-stage systematic sampling of addresses (i.e. clusters). Currently, variance estimation is performed via Taylor linearisation (in STATA). However, jackknife linearisation (in STATA) is suggested by Holmes and Skinner (2000)

(Ref: Variance estimators, ONS, UK, presented in TF-VE)

([14]) However even in this case one has to be careful with the particular assumptions and conventions that each software adopts.

— In the German MicroCensus, one-stage clustered stratified systematic sampling is performed. The assumption of stratified random sampling is made and the corresponding closed-form formulae are used (with no post-stratification), where the used values are aggregated to the level of clusters. It is stated that the omission of systematic aspect leads to an overestimation of the variance.

(Ref: Practice of variance estimation at the German Federal Statistical Office, presented in the 2nd Meeting of TF-VE, Doc. Eurostat/A4/Quality/Variance estimation/01/GFSO).

5.1.2 Multi-stage designs

In multi-stage designs we have more than one component of sampling variability to estimate. More particularly, each stage of the design induces an additional variance component to the total variation of any statistic.

Analytic closed-form formulae can be derived in few, rather simple cases (for example with no post-weighting or, only, complete post-stratification, and simple random sampling at each stage, stratified or not). Generally speaking, however, the ad-hoc development of such formulae is not always easy, or even feasible, so the use of approximate methods is suggested (Taylor linearisation or replication methods). In addition, almost always, advanced software is coupled with complete estimation techniques.

Based on the choice of the most appropriate variance estimation methods, the other characteristics of the sampling design under study and the presentation of software (Table 1) one may proceed to the implementation of variance estimation with the software that most appropriately satisfies his/her requirements.

The simplest case of multi-stage designs is that of ‘two-stage’ designs. In the sequel we discuss, specifically, some designs of this category, commonly used in practice, in household and business surveys. They are usually stratified, while the ultimate sampling unit (SSU in two-stage designs) could be either clusters or elements.

- ***2-stage sampling, with simple random sampling in both stages***

— In this case, jackknife linearisation (performed, for example, in STATA) is an efficient approximate solution, since it holds the nice properties of jackknife variance estimator while it is less time-consuming.

— Of course, another solution (though not always feasible in practice) that could be used is the development of analytic (approximate, usually) formulae.

✍ **Examples**

— In the Austrian Labour Force Survey, where a different sampling design is deployed in the two strata of the population (simple random sampling in one stratum and 2-stage random sampling in the other), the estimation of variance is based on a closed-form approximation formula. According to this formula, the variance consists of two components, representing the variance in each of the two strata. Though this formula incorporates the effects of the particular sampling design (the different sampling designs in the two strata, the clustering in one of the strata and so on) it cannot capture the effect of weighting that is performed on LFS data (using incomplete post-stratification methodology).

— In 'Expenditure and food survey' of ONS, where two-stage stratified random sampling of addresses (i.e. clusters) is performed, Jackknife linearisation (in STATA) is used for variance estimation.

(Ref: Variance estimators, ONS, UK, presented in TF-VE)

- ***2-stage sampling, with systematic random sampling in one or both of the stages***

As we have previously mentioned, there is no analytic unbiased variance estimation in systematic sampling.

— So, a possible approach one could use is to ignore the issue of systematic choice and follow the suggested approach for the above case of two-stage random sampling. The simplification of assuming simple random sampling instead of systematic sampling leads to biased though acceptable variance estimates, as long as the ordering of sampling units has been made according to an influential characteristic of them (leading thus to heterogeneous samples). Moreover, according to Särndal et al. (1992), in multi-stage samples where systematic sampling is deployed in the final stage, the bias is not as serious as one might expect.

— Alternatively, in order to fully incorporate the features of the survey into the variance estimation, one could proceed to the performance of a replication technique. A suggested approach, here, is the balanced half-sample method, which is able to incorporate in the estimation of variance al the features of the sampling design (the stratification as well as the two-stages deployed). The cost for this is the additional computational burden imposed. In large-scale surveys, that are periodically performed, this could turn out to be a prohibitive factor.

— A quicker, but efficient method that could, also, be possibly used is the jackknife linearisation method, which has been proven to be efficient for, the assumed, multi-stage stratified sampling of UK. However, in such a case detailed information on the sorting procedure of the population is required, in order to appropriately group the data.

✍ **Examples**

— In the German LFS (where stratified random sampling is deployed at the 1st stage and systematic clustered at the 2nd), standard errors are calculated in the level of clusters (districts) according to the simple formula for stratified random sampling. In such a case, the effect of post-stratification, non-response weighting and systematic sampling is not taken into account. So, variance estimates are expected to overestimate variance.

— In the Spanish LFS (with the same sample design as Germany), on the contrary, variance estimation is performed via a replication approach, which may adequately take into account both the sampling design and the weighting procedure used. More particularly, balanced repeated replication is deployed.

— The sampling design used in the Irish LFS is two-stage stratified sampling. Systematic sampling is deployed in both stages of the sampling. Post-stratification is also utilized. In order to derive valid variance estimators they adjust the naive estimations derived assuming simple random sampling by an estimated design effect (‘deft’). Different values of ‘deft’ are suggested by Steel (1997) depending on the nature of the phenomenon (variable) under study.

The method of adjustment of variance estimation by ‘deft’ factor is a really cost efficient and straightforward method and it is, undeniably, preferable over the ‘simple random sampling’ simplification. However, it is rather restrictive, in the sense that different factors need to be applied for different measures of interest, which means added burden of work in case a new interest arises. Moreover, in case of repeated surveys on dynamic phenomenon the ‘deft’ factors may need revision and possibly adjustment from time to time.

- ***2-stage sampling, with probability proportional-to-size sampling (in one or both of the stages)***

Generally speaking, probability proportional-to-size sampling is related to complicated inclusion probabilities.

— So, a suggested approach is the use of an appropriate replication method (possibly jackknife). In case of clustered sampling, jackknife is the most preferred replication method.

— Alternatively, GSSE is the only software that can incorporate ppS sampling (CLAN can do only approximate it).

— Following the theoretical framework developed in Belgian LFS one could develop closed-form formulae for the calculation of variances based on inclusion probabilities.

— The option of simplifying assumptions of the sampling design is also present here.

✍ **Examples**

— In the framework of Belgian LFS (with two-stage stratified sampling, 1st stage ppS, 2nd stage systematic sampling of clusters) closed-form formulae have been developed based on (approximate) first and second order inclusion probabilities both for clusters and elements. Numerical techniques, though, are needed for the calculation of inclusion probabilities (the issue of systematic choice of clusters in the 2nd stage is ignored and simple random sampling is assumed, instead).

(Ref: Internal document of statistics Belgium)

— In the Italian LFS (where two-stage sampling design is deployed with stratified πpS and simple random sampling at 1st and 2nd stage, respectively) the calculation of standards errors is performed via GSSE, which is based on the Taylor linearisation method.

— In the Portugal LFS (with stratified πpS and simple πpS sampling at 1st and 2nd stage, respectively), the variance estimation of statistics such as totals is performed in the SAS macro Caljack. Caljack estimates variances based on the jackknife method.

5.2 Incorporation of imputation in variance estimation

5.2.1 General comments

In section 2.3 we have dealt, in a general context, with the impact of imputation in variance as well as in the procedures themselves for variance estimation. Some generic methods for the incorporation of imputation into variance estimation have been mentioned therein (analytic, resampling and repeated imputation methods). These methods need to be adjusted for every specific imputation technique deployed.

As we have previously mentioned, the two potential factors of underestimation of variance (when estimating variance with imputed data) are the following:

— The data set seems to be complete and the factor *1/n* in the formulae is incorrect.

— We ignore the additional variance that is imposed by the imputation.

These issues are illustrated in the following basic example.

Let's assume that we have a simple random sample of size *n*. For a specific question/ variable, there are only *m* respondents and *n-m* imputed data. The response model is that missing data miss at random.

If we neglect the finite population correction, the variance of the mean of *y* is equal to σ^2/m. However the estimator of the variance should be, apparently:

$$\hat{V}_{app} = \frac{1}{n}\frac{\sum(y_k - \bar{y})^2}{n-1} = \frac{1}{n}\hat{\sigma}^2$$

If σ^2 is estimated correctly, we have to make a correction by a factor *n/m*.

Suppose now, that imputation is performed by imputing to the non-respondents the mean of the respondents. In $\sum_s (y_k - \bar{y})^2$ there are *n-m* terms equal to 0, and we have to multiply the quantity by *(n-1)/(m-1)* to get a correct estimate of σ^2. Finally, we have to correct the naive (and given by all standard software!) estimator by a factor approximately equal to $(n/m)^2$! If the response rate is 50%, the factor of underestimation is 4! This is the worst case.

The best case is when the data are imputed by hot–deck, since, in this case, the estimation of σ^2 is correct, and the underestimation is *n/m* only.

Intermediate cases arise when imputation is computed by using a prediction model for the imputed data, with again two cases:

1. The imputed value is the predictor.
2. The imputed value is the predictor plus a 'simulated' residual in order to respect the distribution of the *y* values.

The example can be extended to a more general sampling plan and to a more general estimator.

Different imputation techniques may ask for totally different variance estimation procedures. So, we will not proceed in detail into such a description. However, in the framework of multiple imputation approach a common strategy for variance estimation can be adopted, as discussed in the subsequent section. Furthermore, for reasons of illustration, we present an analytic technique that has been developed for the incorporation of imputation in variance estimation of CVTS2 (Quantos 2001).

5.2.2 Multiple imputation

In multiple imputation, Rubin (1987), each missing value is replaced, instead of a single value, with a set of plausible values that represent the uncertainty about the right value to impute.

So, standard statistical procedures can be implemented, separately, in each one of these multiply-imputed data sets. Subsequently, one may combine the results of these analyses and come up with single, common, estimates.

It is important to note that the process of combining results from different imputed data sets is essentially the same, regardless of the type of statistical analysis. This leads to valid statistical inferences that properly adjust for the non-response even in complicated cases (Herzog and Rubin, 1985).

Summarising, as mentioned in Yuan (2000), multiple imputation inference involves three distinct phases:

- The missing data are filled in *m* times to generate *m* complete data sets.
- The *m* complete data sets are analysed by using standard procedures.
- The results from the *m* complete data sets are combined for the inference.

Multiple imputation approach, its theoretical justification and its use for statistical inference in practice is discussed in detail in Herzog and Rubin (1985). In this case the incorporation of imputation (more precisely the variance induced to the estimates from the imputation procedure) can be easily derived based on the variability of the estimates among the multiply imputed data sets.

More particularly, Herzog and Rubin (1985) show that the resulting variance estimation (when multiple imputation is deployed) can be estimated as the sum of the following two components:

1. The average variance of estimation given one set of imputed values
2. The variance of estimates across the multiple imputations.

In Luzi and Seeber (2000) one may find a simulation study evaluating several imputation methods (including multiple imputation) and corresponding software in terms of precision, i.e. variance estimation.

5.2.3 A case-study

In the case of CVTS2, conducted in 2000, general guidelines have been provided for the imputation method deployed in each country, discriminating between quantitative and qualitative variables (Eurostat 2000a,b).

Some critical features of this survey are:

— Sampling and original processing of the data (i.e. imputation) are performed at national levels.

— Datasets at micro level are transferred to Eurostat, where analysis (including variance estimation) is made.

— No flagging of the imputed values was available (transmitted to Eurostat).

— Different imputation has been deployed for quantitative and qualitative variables.

5.2.3.1 *Qualitative variables*

In the case of qualitative variables sequential hot-deck imputation was performed, while no-flagging of imputed values was available in the final datasets transmitted to Eurostat.

So the lack of precise information and the diversity of the methods in addition to the deficiency of appropriate software impose to apply some rough approximation. Taking into account the fact that the suggested imputation method is near to the 'best case' described above, we propose the *two-step procedure*:

1. compute the variance as if there were no imputation; and
2. multiply this estimate by the **inverse of the response rate** (correction factor).

For example if there are 4 000 units in the sample but only 3 000 have responded to the specific question then the variance has to be increased by 1.33 (4 000/3 000) and the confidence intervals would be increased by the square root of this factor. Of course this adjustment holds for simple random sampling and we have to use it under the assumption that the effect is uniform across all strata.

Nevertheless, this correction may be considered as very conservative and thus it can be argued that it should be used as an upper limit only. The correction factor assumes that imputation does not add anything to the precision of the estimation. However, if the imputation model is good we should expect that some values must have been correctly imputed. Thus, while it is unreasonable to expect that all the values are going to be perfectly predicted it is also unreasonable to expect that none will be close to the true value if the imputation model is correct.

5.2.3.2 *Quantitative variables*

For quantitative variables, the imputation method falls under the general setting of 'ratio imputation'. According to this approach, a ratio $\hat{r}_r = \dfrac{\hat{t}_r^y}{\hat{t}_r^x}$ is computed on the respondent sample (r). Then, for k in nr (set of non-respondents), the value $y_k = \hat{r}_r x_k$ is imputed. Note that for the purposes of the imputation procedure of variable Y, it is assumed that there is no item non-response for variable X, that means that either there is, indeed, no item non-response or missing values have already been imputed in a previous step of the analysis.

That is, the finally derived variable Y has values:

$$y_k^* = \begin{cases} y_k & \text{if } k \text{ belongs to the respondents set } (r) \\ \hat{r}_r * x_k & \text{if } k \text{ belongs to the non - respondents set } (nr) \end{cases}$$

Starting from this fact, there are many ways to tackle the problem of variance estimation. The more simple and general approach for the variance estimation of the total of variable, let's say, Y is as follows:

We have $\hat{t}^y = \hat{t}_s^x \hat{r}_r$ and therefore $V(\hat{t}^y) \cong \left(t^x\right)^2 V(\hat{r}_r) + r^2\, V(\hat{t}_s^x)$

So, in order to properly assess the variance of variable we need to get point and variance estimates for $\hat{t}^x$ and $\hat{r}_r = \hat{t}_r^y / \hat{t}_r^x$

- Appropriate point estimates for t^x and r can be easily derived ignoring the issue of imputation using standard software.
- The variance estimation of $\hat{t}_s^x$ is also derived rather straightforward since:
  - in the case that it is, actually, a variable with no item non-response, the variance estimation can be easily derived via several of the software of Table 1;
  - in case that imputation has already been performed to substitute for missing values, the correct variance estimation will have already been computed in a previous step and thus it will be readily available for this step.
- Thirdly, we have to calculate $\hat{V}(\hat{r}_r)$, that is the variance estimation of a ratio with missing values.

If the imputed data were flagged, there would be a rigorous solution as the ones mentioned in Deville and Särndal (1994) and Lee et al. (1994).

However, in CVTS2, the datasets transmitted to Eurostat do not have imputed values flagged. Thus, we have to use a 'rough correction' method. The methodology proposed here is a **four-step procedure** as follows:

1. We construct the 'residuals' $y_i - \hat{r}_r x_i$ and calculate their variance as if all data were observed, and not imputed.
2. Residuals' variance is inflated by the square of the inverse of the response rate in order to provide an adequate variance estimation for $\hat{r}_r$.

   (*The rationale for this is that the 'residuals' are equal to zero for imputed data, and the estimator behave like the mean imputation estimator as previously described).*

   - Finally, using the estimates of the components in the three steps above, we can derive an appropriate estimator for the measure that originally concerned us, that is variance estimation of Y, through the formula:

$$\hat{V}(\hat{t}^{\,y}) \cong \left(\hat{t}_s^{\,x}\right)^2 \hat{V}(\hat{r}_r) + \hat{r}_r^2 \, \hat{V}(\hat{t}_s^{\,x})$$

This proposed method can be further elaborated for the derivation of variance estimators, taking into account imputation, of other statistics such as means, percentages or ratios. The case of variance estimation of these statistics, and in particular of ratios, raises some difficulties, since we have somewhere to evaluate the covariance between two estimators based on different samples overlapping at random.

5.3 Special issues in variance estimation

5.3.1 Variance estimation in the presence of outliers

A problem, frequently occurring in business surveys, in particular, is that of outlier observations, that is observations with 'unexpectedly' extreme (large or small) values in one or more of the measured variables.

The issue of outliers in surveys and possible strategies for dealing with them are discussed analytically by Chambers (1998).

According to them, there are three main approaches to dealing with sample outliers.

- Deletion of outliers from the sample

This method, although commonly used in practice, is the weakest with respect to its theoretical justification.

- Assignment of weight equal to 1 to the outlier observations

This type of weighting implies that outliers are, essentially, unique in the population (i.e. they represent only themselves).

- Modification of outliers so as to reduce their impact on the sample estimates

In this approach, known an winsorisation, we retain the original weight of the value, as calculated by the corresponding weighting method, but we try to modify the extreme value itself, so as to diminish its impact on the estimates. One-sided winsorisation, applicable to strictly positive variables, commonly met in business or household surveys, has a sound theoretical background, since the modifications are model-based allowing thus statistical inference.

Currently, there is a great degree of heterogeneity among the methods for outlier treatment used at surveys, while differentiations exist even to the definition of outliers themselves.

Outlier observations, having large deviations from the remaining cases, tend to increase the estimates of variance. Each one of the aforementioned approaches for dealing with outliers, leads to a stabilisation of the variance estimates introducing though a component of bias in the estimation, since standard variance estimation methods do not take into account the outlier treatment applied (underestimating, thus, true variance). Actually, there exists a trade-off among these two issues (decrease of variance and increase of bias). Deletion of outliers and assignment of weights equal to one, stabilize variances but on the cost of a potentiality large bias. Winsorisation methods may control more efficient this trade-off of bias and variance via the use of sophisticated models.

5.3.2 Variance estimation within domains

In most surveys, our interest is focused not only on the estimates for the whole sample population of study but also for specific subpopulations, usually called domains. Domains may be overlapping or not, complementary (covering the whole population) or not. Often domains cut across stratum boundaries and are referred to as 'cross-classes'.

For example, in household surveys we might be interested in inference distinguishing among families of different size (with 1, 2 etc members) or of different socio-economic class. Furthermore, in cases of out-of date sampling frames, where the original stratum classification of observations deviates from the actual current status of sampled units, domain estimates are essentially needed for an analysis based on the observed classification of units. This is a phenomenon often occurring in business surveys.

So since, usually, the formation of these subpopulations of interest is unrelated to the sample design, the sample sizes for the subpopulations are random variables, inducing an additional component of variability into the domain estimates. This distinguishes domain estimates from, for example, estimates within strata.

The stochastic nature of the size of the domains, makes invalid the mere application of any estimation method (if we regard as fixed the sample sizes of the domains then we tend to underestimate the variances, since we ignore a component of additional variability).

The key idea here is that if we transform the variable of interest as:

$y^*_{hi} = \begin{cases} y_{hi}, \text{if i} \in \text{domain D} \\ 0, \text{otherwise} \end{cases}$ then, according to the design-based approach to inference (Chambers, 1998), the application of standard methods to the transformed variable y^* provides us with the appropriate estimates.

Many of the previously reviewed software facilitate the estimation of several quantities for different domains of the population.

Analytic formulae, based on Taylor linearisation method can be provided for variance estimation of simple statistics within domains (Chambers, 1998). Alternatively, as noted in Brick et al. (2000), replication methods for variance estimation are also well suited for the analysis of domains. In this case, where the replicate weights contain all the required information for variance estimation, one may work only with the subset file containing the domain of interest and the replicate weights.

5.3.3 Variance estimation with one-unit per stratum

A problematic situation for the procedure of variance estimation arises when in a realized sample we have only one unit per stratum (in all or some of the strata of the sample). This situation may be confronted when we have a very refined stratification and:

- each stratum has sample size greater than one, but only one responding unit exists;
- the design of the sampling, itself imposes a single-unit per stratum.

In any of these two cases, it is impossible to directly calculate variances with one sampled unit per stratum. The most common strategy, suggested in the literature, for dealing with this problem is the collapsed stratum technique. The issue of collapsing strata for variance estimation with one unit per stratum is discussed in Cochran (1977) and Sarndal et al. (1992) (Wolter, 1985, deals with the case of multi-stage designs with one PSU per stratum).

Note, however, that a critical point for the effectiveness of this method relates to the origin of the one-unit per stratum situation. If the problem comes from non-response, it could mean that the chosen estimator is dramatically bad (with very low reliability).

In the framework of this technique one essentially proceeds to the merging (collapsing) of two or more strata in order to calculate variances in groups with more than one units, using the most appropriate among the aforementioned techniques. Generally speaking, this collapsing of groups is biased and leads to an overestimation of the variance. In order to reduce the bias, the strata to be groups should be chosen to be as similar as possible, since the bias depends on the difference of the means between the unit strata used for collapse. However, what is equally important is to select the complementary stratum carefully based on a priori information. That is, the allocation of groups should be based on the population characteristics and not on the sampling ones.

The main steps for collapsing stratum technique, are illustrated below for the simple case where we have L (even number of) strata, each consisting of a single unit, and we intend to collapse them in G groups of two (G=L/2). The steps that could be followed are

i) Identification of the strata that are going to be collapsed

This type of grouping is required in order to diminish the overestimation that is inherent to the ‘collapsed-stratum’ variance estimation.

ii) The corresponding variance of the total $\hat{t}^{y}$ can, indeed, be estimated as the sum of G components as follows:

$$\hat{t}^{y} = \sum_{g=1}^{G} \hat{t}_{g}^{y} \text{ , where } \hat{t}_{g}^{y} = \hat{t}_{g1}^{y} + \hat{t}_{g2}^{y}$$

So, $\hat{V}\left(\hat{t}^{y}\right) = \sum_{g=1}^{G} \hat{V}\left(\hat{t}_{g}^{y}\right) = \sum_{g=1}^{G} \left(\hat{t}_{g1}^{y} - \hat{t}_{g2}^{y}\right)^{2}$

Further modifications are required for the above formulae in case that not all strata are single-unit and/or there are strata with different number of units. These adjustments are needed due to the fact that the units that constitute the collapsed strata have unequal probabilities of selection (see for example Kish, 1965).

It worth noting that apart from the aforementioned analytic formulae, appropriate replication methods for variance estimation can accommodate the collapsing of strata such as BRR or JK2 (Morganstein, 1998).

5.3.4 Non-parametric confidence interval for median

Apart from totals, means, proportions or ratios there are more complicated statistics for which one may be interested. Such an example is the median of a variable, which is essentially a function of the probability distribution function of the variable. Median and other percentiles of a distribution are particularly useful in cases of skew variables, as often met in business (e.g. salaries) and household surveys (e.g. incomes). In these cases more elaborate techniques are required for the calculation of corresponding variances.

Särndal et al. (1992) present a non-parametric technique for the calculation of confidence interval for medians (the variance can be derived indirectly).

The main steps of this approach are as follows:

- Produce an estimate of the probability distribution, let's say $\hat{F}$, using the philosophy of ratio estimation.
- Proceed to construction of variance estimation formulae for any point of the estimated probability distribution (i.e. for any probability). The assessment of those variances is based on the fact that we are dealing with ratio estimates.
- Inversing the estimated distribution $\hat{F}$, estimate the median as $\hat{M}$ (for p=0.5).
- Approximate a confidence interval for the estimated median by inversing a confidence interval for the 50% probability.

This procedure can be applied to sample designs more complicated that simple or stratified random sampling, if appropriately incorporate the inclusion probabilities in the estimation of $\hat{F}$

The main advantages of such a non-parametric approach are:

- no model assumptions are made on the population, since it is based on the design-based approach to variance estimation;
- it makes use of closed-form formulae for variance estimation and thus it can be extended to complex survey designs and large samples;
- it is less time-computing demanding than resampling methods.

However, some restrictions are:

- The populational probability distribution function of the variable under investigation should be continuous. In discrete cases the estimated confidence interval has larger coverage than the nominal one. This causes problem for finite populations (since in these cases the distribution is essentially discrete), however, the overcoverage can be considered to be negligible if the population is large enough.
- As we previously saw, the original variance estimation is made for the distribution function, and an inverse is needed in order to derive an estimate for the median. In this procedure approximations are deployed.

This technique has been adopted by Graf (2002) for the derivation of confidence interval for the median in the Swiss earnings structure survey 2000 (SESS-2000). The author reached a number of practical findings:

- if the empirical distribution function is bimodal, in which case the median is an inappropriate measure anyway, the confidence intervals tend to be rather large;
- on the contrary the narrowest confidence intervals are observed when the empirical distribution function is steep around the median;
- observations with very large weights induce instability in the median estimation. The empirical distribution function cannot be regarded as continuous, making, thus, necessary the elimination of largest weights.

5.3.5 Field substitution

5.3.5.1 *Main characteristics of field substitution*

An issue of particular interest in the context of variance estimation is the practice of ‘field substitution’ commonly used in household sample surveys.

Field substitution (for brevity, in the sequel called merely substitution) is an alternative technique, to imputation or weighting, in order to deal with unit non-response. That is, instead of imputing data from respondents or adjusting the weights of the respondents, population units not initially selected for the sample are used to replace eligible sample units that do not participate in the survey.

The most appealing feature of field substitution is that it **preserves the optimal structure of the survey**. By substitution, we are able to keep fixed, as originally planned, the size and the allocation of total samples within stages, strata and intermediate sampling units. This is particularly appealing in complex sample designs, especially when small strata or clusters are deployed. For example in case of stratified sampling with two sampling units (elements or clusters) per stratum, substitution alleviates the problem of single units per stratum that may occur due to non-response and which calls for other techniques, such as collapsing strata, in order to deal with the corresponding variance estimation.

Some other advantages of the field substitution are the following:

- control of the sample size;
- Removal (at least, partial) of the non-response bias;
- simplicity for the users.

The disadvantages of the field substitution procedure are, primarily, related to the following:

- provision of illusion that non-response has been removed;
- non-response rate tends to increase;
- field work is prolonged.

5.3.5.2 *Implication of field substitution in variance estimation*

The main discussion on field substitution is related to its effect on the bias that induces on the estimates. However, ignoring any possible bias, i.e. assuming that non-responses are 'missing at random', we may study its impact on the variance estimation.

Generally speaking, comparing variance from a sample, in which substitution has been deployed, with a sample of equal size with no non-response, no significant differences arise (Vehovar, 1999). Large difference may occur in rare cases, such as when there are strong dissimilarities among secondary respondents and secondary non-respondents. In complex sample designs as the ones most often used in national LFS (incorporating stratification and clustering) the impact of field substitution in variance estimation tends to be negligible.

Vehovar (1999) deals, in detail, with the case of a two-stage cluster sampling. In this case, benefit appears (compared to weighting adjustments) only in very specific circumstances (rarely met in practice) with respect to the size of the clusters and the intra-class correlation. For more than two stages the benefit is negligible, since substitution is performed at the last stage clusters, and the corresponding component of the within cluster variability represents an even smaller part of the sampling variance. On the other hand substitution at the level of primary sampling unit is even more impractical as the cluster size would be too big to allow any gains in variance.

Similar results apply to stratification, where the gains in precision are similar to the case of proportionate sample compared with post stratification (Cochran, 1977). However, in practice, the strata are usually too large in order to benefit from field substitution.

5.3.5.3 *Practical guidelines for the use of field substitution*

As mentioned in Chapman (1983) field substitution is more appropriate for **surveys that involve an extensively stratified design, with a relative small sample size**. In such cases, field substitution will tend to provide a better treatment of the non-response bias, compared to weighting or imputation. This is primarily due to the following reasons:

- in strata with small extent it is easier to find a substitute unit with quite similar characteristics to a non-respondent;
- on the contrary, in order to perform post-stratification adjustment, wider classes need to be developed (to avoid producing non-trivial increase in variance estimates), loosing thus the relative advantage over substitution.

More generally, field substitution could be considered as an alternative method for dealing with non-response, in the following cases:

- where there is a strong need for a self-weighted sample. In such a case the following conditions must additionally hold:
    - there are no other theoretical reasons for weighting,
    - the substitution can remove the non-response bias, at least to the extent of the alternative procedures;

- it is possible, due to high non-response and a small 'take' per cluster (or stratum), a considerable number of clusters (or strata) to have no (or only 1) observations([15]);

- there exists a potential advantage of improved precision, with respect the mean squared error (for a more elaborate discussion on this issue one may refer to Vehovar, 1999).

On the contrary, field substitution is **not** an appropriate procedure for large probability samples where at least one of the following features is valid:

- short time available for field operations;
- evidence of a strong bias induced by the substitution procedure itself;
- weak (or expensive) control over field work procedures.

Finally, we should point out that there is no general theoretical framework supporting or rejecting the use of field substitution. The efficiency of its use seems to be a matter of empirical-only evaluation. That is, in order to evaluate its performance once should compare it with alternative adjustment methods (such as imputation or post-weighting).

To conclude with, some conventions, in cases where 'field substitution' is used, are the following:

- identification of the data records that are obtained from substitute units should be kept;
- the level of substitution should be reported;
- substitutes should be treated as non-response cases in the calculation of the survey response rate.

5.4 Calculation of coefficients of variation

Apart from the variances themselves, another very useful measure, indicative of the quality of any estimate, is the coefficient of variation (relative standard error). Actually, coefficients of variation are of the most popular quality indicators of statistics. Coefficients of variations can be computed either at national level or, even, at EU level. Some generic instructions for its derivation are provided in the paragraphs to follow.

5.4.1 National level

For any statistic θ, estimated by $\hat{\theta}$ and with estimated variance $\hat{V}(\hat{\theta})$, one can estimate([16]) its coefficient of variation as:

([15]) This situation may be encountered in surveys of institutions (stores, schools) and in specific household surveys (such as the 'household budget survey').

([16]) The symbol CV(.) indicates the estimator of the true coefficient variation of any statistic.

$$CV\left(\hat{\theta}\right)=\frac{\sqrt{\hat{V}\left(\hat{\theta}\right)}}{\hat{\theta}}$$

Such a measure can be calculated either for the whole population under study or for separate domains of it. In the latter case, the same formula, as above, applies as well, where point and variance estimates are calculated for each domain separately. For example, the coefficient of variation of statistic *S* for domain *d* is calculated as:

$$CV_d\left(\hat{\theta}\right)=\frac{\sqrt{\hat{V}_d\left(\hat{\theta}\right)}}{\hat{\theta}_d}$$

where $\hat{\theta}_d$, $\hat{V}_d\left(\hat{\theta}\right)$ are the point and variance estimates of statistic θ for domain *d*, respectively.

5.4.2 EU Level

When it comes to EU level, the above formula holds as well. The critical point here is to obtain a common point and variance estimate for the whole EU (combining the national estimates).

This is not a hard task though, since we can treat each country as a separate stratum with independent sampling, whose union constitutes the whole of EU. So, the logic that is used in the stratification can be applied here as well, i.e. point and variance estimates are essentially (weighted) averages of the corresponding estimates of all strata.

- For the case of a *total*, we have that:

$$\hat{t}_{EU}=\sum_i \hat{t}_i\ ,\ \hat{V}\left(\hat{t}_{EU}\right)=\sum_i \hat{V}\left(\hat{t}_i\right)\Rightarrow$$

$$CV\left(\hat{t}_{EU}\right)=\frac{\sqrt{\hat{V}\left(\hat{t}_{EU}\right)}}{\hat{t}_{EU}}=\frac{\sqrt{\sum_i \hat{V}\left(\hat{t}_i\right)}}{\sum_i \hat{t}_i}.$$

(subscript *i* refers to the national estimates of Member States $\{i\}$, and *EU* to the aggregate estimate at EU level)

- For the case of a *mean*, we have that:

$$\hat{\mu}_{EU}=\sum_i \frac{N_i}{N_{EU}}\hat{\mu}_i\ ,\ \hat{V}\left(\hat{\mu}_{EU}\right)=\sum_i \frac{N_i^2}{N_{EU}^2}\hat{V}\left(t_i\right)\Rightarrow$$

$$CV\left(\hat{\mu}_{EU}\right)=\frac{\sqrt{\hat{V}\left(\hat{\mu}_{EU}\right)}}{\hat{\mu}_{EU}}=\frac{\sqrt{\sum_i N_i^2\hat{V}\left(\hat{\mu}_i\right)}}{\sum_i N_i\hat{\mu}_i}.$$

Note: The above formulae for the case of mean hold if the sizes of the populations (N_i) are considered to be known. Alternatively (if they are considered unknown), one should refer to the case of CVs for ratios, further discussed below.

- For the case of a *ratio*:

In the case of ratios, and more generally of non-linear statistics, the formula is more complicated, mainly due to the fact that the variance at EU level cannot be decomposed into a weighted sum of the national variances. So, here we have:

$$\hat{r}_{EU} = \frac{\hat{t}^y_{EU}}{\hat{t}^z_{EU}} = \frac{\sum_i \hat{t}^y_i}{\sum_i \hat{t}^z_i}$$

$$\hat{V}(\hat{r}_{EU}) \cong \frac{1}{\left(\hat{t}^z_{EU}\right)^2}\left[\hat{V}\left(\hat{t}^y_{EU}\right) + \hat{r}^2_{EU} \cdot \hat{V}\left(\hat{t}^z_{EU}\right) - 2 \cdot \hat{r}_{EU} \cdot \hat{C}\left(\hat{t}^y_{EU}, \hat{t}^z_{EU}\right)\right]$$

$$\hat{C}\left(\hat{t}^y_{EU}, \hat{t}^z_{EU}\right) = \sum_i \left\{\hat{C}\left(\hat{t}^y_i, \hat{t}^z_i\right)\right\}$$

Summing up, the required information (at national level) for the calculation of EU CVs, depending on the type of statistic is illustrated in the table that follows. Note, that it is preferable that all estimators (point and, especially, variance) are calculated using the same method (so that they have the same properties and level of precision, accuracy).

Table 2: Information required for EU CV's

Statistic	*Required information at national level*
Total	Total estimate; Variance estimate of total
Mean	Total estimate; Variance estimate of total; Population size
Ratio	Total estimate (numerator, denominator); Variance estimate of total (numerator, denominator); Covariance of totals (numerator-denominator)

5.4.2.1 *National contribution to EU coefficient of variation*

In case that we are mainly interested in EU CVs and not national ones, an alternative measure that could be presented, accompanying the EU CVs is the contribution of each country to this total figure. More precisely, the contribution of MS *i* to the EU CV (CT_i) is calculated as:

$$CT_i = 100 \cdot \frac{CV_{EU} - CV_{EU-\{i\}}}{CV_{EU-\{i\}}}$$

where CV_{EU} is the coefficient of variation of statistic S calculated at EU level, and

$CV_{EU-\{i\}}$ is the corresponding coefficient obtained if country *i* is ignored in the estimation of CV.

A positive contribution indicates that MS *i* tends to increase the variability of statistic S at European level, while the opposite (decrease of EU variability) is suggested by a negative value of contribution.

A case study of the implementation of CVs in structural business statistics (at European Level) is provided in ‘Coefficient of Variations in Structural Business Statistics’ presented at the 2nd meeting of TF-VE.

6. CONCLUDING REMARKS

In this report we have tackled the issue of variance estimation in the framework of several sampling designs and estimation procedures that are currently used in EU countries, covering both household and business surveys. Variance estimation is a crucial issue in the assessment of the survey results.

As we have previously mentioned, the choice of an appropriate variance estimator primarily depends on:

- the type of the estimator itself;
- the type of adjustments (weighting procedures) performed;
- the underlying sampling design of the survey.

These factors complicate the straightforward estimation of variance and a number of more advanced variance estimation methods have been developed. Thus, the literature on variance estimation is rich, however no clear guidelines exist, for each case more than one techniques could be used, each one of which suffering from some drawbacks and offering some other advantages, at the same time.

In Table 3 that follows, we summarise the several alternative variance estimation methods and practices that have been reviewed in section 5.1 for business and household surveys conducted in EU countries. To enhance the readability of the table we note the following: the first column ‘Sampling design’ describes the several sampling designs and situations that have been met in practice in European level. In the second column (‘Current practices’) examples of specific surveys and the methods used in practice of variance estimation are presented. For each case, corresponding variance estimation approaches are suggested in the final column of the table (‘Suggested methods’). Distinction is made between business and household surveys. For a further discussion on these issues of current practices and suggested methods one may refer to section 5.1.

In the final choice of the most appropriate variance estimation method, additional factors need to be taken into account (when one has to choose among more than one valid appropriate estimator) such as:

- properties of the resulting variance estimator (bias, mean square error, coverage of confidence interval);
- timeliness (ad hoc calculation vs. general software);
- operation convenience (simplicity);
- information required for calculation (and availability of such information on confidentially protected files;
- other administrative details.

The final choice is, essentially, a trade-off among the above criteria.

Table 3: Comparative presentation of variance estimation methods for business and household surveys

Sampling designs	Current practices (business surveys)	Suggested methods
One-stage designs		
Sampling of elements		
Simple random sampling	○ CVTS2 at Eurostat (*Stratified*): no estimation	• Taylor linearisation
Systematic sampling	○ Trade statistics in Germany (GFSO) (*Stratified*): Taylor's linearisation	• Jackknife linearisation • Replication methods • Pseudo-stratification

Sampling designs	Current practices (household surveys)	Suggested methods
One-stage designs		
Sampling of elements		
Simple random sampling	○ LFS Luxembourg: closed form formulae	• Closed form formulae (exact for linear statistics, Taylor's linearisation for non-linear statistics)
Systematic sampling	○ LFS of Finland: Taylor's linearisation ○ LFS of Sweden (*Stratified*): Taylor's linearisation	• Jackknife linearisation • Replication methods • Pseudo-stratification
Probability proportional-to-size sampling		• Replication methods • Jackknife linearisation • Use of GSSE
Sampling of clusters	*Clustered systematic (stratified):* ○ LFS of UK — Taylor's linearisation ○ German MicroCensus: closed-form Formulae	• Taylor's linearisation • Jackknife linearisation • Replication methods (jackknife)
Multi-Stage Designs		
2-stage sampling, with simple random sampling in both stages	○ LFS of Austria: closed-form Formulae ○ Expenditure and Food Survey of UK: jackknife linearisation	• Jackknife linearisation • Closed form formulae (approximate)
2-stage sampling, with systematic random sampling in one or both of the stages	○ In general effect of systematic selection is ignored ○ LFS of Germany: simplified closed-form formulae ○ LFS of Spain: replication method (BRR) ○ LFS in Ireland: use of design effect	• Replication methods (BRR) • Jackknife linearisation

Sampling designs	Current practices (household surveys)	Suggested methods
2-stage sampling, with probability proportional-to-size sampling (in one or both of the stages)	○ LFS of Belgium: closed form formulae (approximate) ○ LFS of Italy: closed form formulae (approximate) ○ LFS of Portugal: replication methods (jackknife)	• Replication methods (jackknife) • Use of GSSE • Closed form formulae (approximate)

7. REFERENCES

Andersson, C. and Nordberg, L. (1998). *A User's guide to CLAN 97*. Statistics Sweden.

Bell, P. (2000). 'Weighting and standard error estimation for ABS household surveys,' prepared for ABS Methodology Advisory Committee.

Berry, B. L. and Baker, A. M. (1968). 'Geographic sampling.' *Spatial analysis*, (Berry, B. L. and Marbles, D. F., editors), Prentice-Hall, Eaglewood Cliffs.

Brick, J. M., Morganstein, D. and Valliant, R. (2000). *Analysis of complex sample data using replication*. Westat

Canty, A. J. and Davison, A. C. (1998). 'Variance estimation for two complex surveys in Switzerland,' *Ecole Polytechnique Federale de Lausanne*.

Canty, A. J. and Davison, A. C. (1999). 'Resampling-based variance estimation for labour force surveys,' *The Statistician*, **48(3)**, pp. 379–391.

Chambers, R. (1998). Probability sampling: Extension, in 'Model Quality Report in Business Statistics', Vol I: *Theory and Methods for Quality Evaluation* eds. Davies, P. and Smith, P.

Chapman, D. (1983). 'The impact of substitutions on survey estimates.' *Incomplete data in sample surveys*, Vol. II, Theory and bibliographies, eds. Madow, W., Olikn, I. and Rubin, D. New York: National Academy of Science, Academic Press, pp. 45–61.

CIRCA (2001) *Practice of variance estimation at the German Federal Statistical Office,* presented in the 2nd Meeting of TF-VE, Doc. Eurostat/A4/Quality/Variance estimation/01/GFSO.

CIRCA (2001) *Practice of variance estimation at the German Federal Statistical Office,* presented in the 2nd Meeting of TF-VE, Doc. Eurostat/A4/Quality/Variance estimation/01/GFSO

CIRCA (2001) *Questionnaire from Statistics Finland*, presented in 3rd Meeting of TF-VE.

CIRCA (2001) *Variance Estimators, ONS, UK*, presented in TF-VE.

CIRCA (2001) *Variance calculation software: evaluation and improvement*, (Supcom 1998 project, Lot 16)

Cochran, W. G. (1977). *Sampling techniques*. J. Wiley, New York.

Deville, J. C. and Särndal, C. E. (1992). 'Calibration estimators in survey sampling,' *Journal of the American Statistical Association*, **87**, pp. 376–382.

Deville, J. C. and Särndal, C. E. (1991). 'Estimation de la variance en présence de données imputées,' *Proceedings of Invited Papers for the 48th Session of the International Statistical Institut*e, Book 2, Subject 17, pp. 3–17.

Deville, J. C. and Särndal, C.E. (1994). 'Variance estimation for the regression imputed Horvitz-Thompson estimator,' *Journal of Official Statistic*s, **10**, pp. 381–394.

Estevao, V., Hidiroglou, M. A. and Särndal, C. E. (1995). 'Methodological Principles for a Generalised Estimation System at Statistics Canada,' *Journal of Official Statistics*, **11(2)**, pp. 181-204.

Eurostat (2000). European Union Manual: Continuing vocational training survey (CVTS2).

Eurostat (2000). Imputations in the CVTS2.

Full, S. (1999). 'Estimating variance due to imputation in ONS business surveys,' presented at the International Conference on Survey Non-Response.

Gagnon, F., Lee, H., Rancourt, E. and Särndal, C. E. (1997). 'Estimating the variance of the generalised regression estimator in the presence of imputation for the generalised estimation system,' *Proceedings of the survey methods section, Statistical Society of Canada*, pp. 151–156.

Graf, M. (2002). 'Nonparametric confidence interval for the median in a stratified two stage cluster survey: The Swiss earnings structure survey 2000 (SESS-2000),' *Statistical Methods Unit, Swiss Federal Statistical Office* (presented at 4th Meeting of Task Force: Variance Estimation, Eurostat).

Herzog, T. N. and Rubin, D. B. (1985). 'Using Multiple Imputations to Handle Nonresponse.' *Incomplete data in sample surveys*, Vol. II, Theory and Bibliographies, eds. Madow, W., Olikn, I. and Rubin, D. New York: National Academy of Science, Academic Press, pp. 209–245.

Holmes, D. G. and Skinner, C. J. (2000). 'Variance estimation for labour force survey estimates of level and change.' *GSS Methodology Series*, No **21**.

Kish, L. (1965). *Survey sampling*. J. Wiley and Sons, New York.

Kish, L. and Frankel, M. (1974). 'Inference from Complex Samples,' *Journal of the Royal Statistical Societ*y, **B36**, pp 1–22.

Kish, L., Namboodiri, N. K. and Pillai, R. K. (1962). 'The ratio bias in surveys,' *Journal of the American Statistical Association*, **57**, pp. 863–876.

Kott, P. S. (1998). 'Using the delete-a-group jackknife variance estimator in practice,' *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 763–768.

Kovar, J. G. and Whitridge, P. J. (1995). 'Imputation of business survey data,' *Business Survey Methods*, eds Cox, B. G., Binder, D. A., Chinnappa, B. N., Christianson, A., Colledge, M. J. and Kott, P. S., pp 403–423. New York. Wiley.

Kovar, J. G., Rao, J. N. K. and Wu, C. F. J. (1988). 'Bootstrap and other methods to measure errors in survey estimates,' *Canadian Journal of Statistic*s, **16**, pp.25–45.

Krewski, D. and Rao, J. N. K. (1981). 'Inference from stratified samples: Properties of the linearisation, jackknife, and balanced repeated replication methods,' *Annals of Statistic*s, **9**, pp. 1010-1019.

Lee, H., Rancourt, E., and Särndal, C. E. (1999). 'Variance estimation from survey data under single imputation,' presented at the *International Conference on Survey Non-response*, to be published in a monograph.

Lee, H., Rancourt, E., Särndal, C. E. (1994). 'Experiment with variance estimation from survey data with imputed values,' Journal of Official Statistics, 10, pp. 231–243.

Lemaitre, G. and Dufour (1987). 'An integrated method for weighting persons and families,' *Survey Methodology*, **13 (2)**, pp. 199–207.

Luzi, O. and Seeber, A. C. (2000). 'Multiple imputation — An experimental application,' prepared for Working Group on the Assessment of Quality on Statistics, Eurostat.

Mohadjer, L., Yansaneh, I. and Brick, M. (1996). 'Application of logistic models to survey data using replication techniques,' *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 462–467.

Montaquila, J. M. and Jernigan, R. W. (1997). 'Variance estimation in the presence of imputed data,' *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 273–278.

Morganstein, D. (1998). 'The replication method for estimating sampling errors,' *International Statistical Seminars*, Eustat.

Patacchia, O. (2000). Imputation — overview of methods with examples of procedures used in Eurostat. Unit A4 (Research and Development, Methodology and Data Analysis) — Eurostat.

Quantos (2001). Report on variance estimation in CVTS2. Lot 4, 2000/S 135-088090 invitation to tender.

Rancourt, E. (1996). 'Issues in the combined use of Statistics Canada's generalised edit and imputation system and generalised estimation system,' *Survey and Statistical Computing: Proceedings of the Second ASC International Conference,* Association for Survey Computing, pp. 185–194.

Rancourt, E. (1998). 'Estimating variance in the presence of imputation: The state of play system,' presented at the *Journees de Methodologie Statistique*, Paris.

Rao, J. N. K. (1990). 'Variance Estimation under Imputation for Missing Data,' Technical report, Statistics Canada, Ottawa.

Rao, J. N. K. (1991). 'Jackknife variance estimation under imputation for missing data,' Technical Report, Statistics Canada, Ottawa.

Rao, J. N. K. (1997). 'Developments in sample survey theory: An appraisal,' *Canadian Journal of Statistics,* **25**, pp. 1–21.

Rao, J. N. K. and Shao, A. J. (1992). 'Jackknife variance estimation with survey data under hot-deck imputation,' *Biometrika*, **79**, pp. 811–822.

Rao, J. N. K. and Shao, J. (1996). 'On balanced half-sample variance estimation in stratified sampling,' *Journal of the American Statistical Association*, **91**, pp. 343–348.

Rao, J. N. K. and Sitter R. R. (1995). 'Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika*, **82**, pp. 453–460.

Rao, J. N. K., Wu, C. F. J. and Yue, K. (1992). 'Some recent work on resampling methods for complex surveys,' *Survey Methodology*, **18**, pp. 209–217.

Rubin, D. B. (1978). ‘Multiple imputation in sample surveys — A phenomenological bayesian approach to non-response,’ *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 20–34.

Rubin, D. B. (1987). *Multiple imputation for non-response in survey*s. John Wiley and Sons, New York.

Särndal C. E., Swensson, B. and Wretman. J. (1992). *Model assisted survey sampling*. Springer-Verlang. New York, Inc.

Särndal, C. E. (1992). ‘Methods for estimating the precision of survey estimates when imputation has been used,’ *Survey Methodology*, **18**, pp. 241–252.

Shao, J. and Sitter, R. R. (1996). ‘Bootstrap for imputed survey data,’ *Journal of the American Statistical Associatio*n, **91**, pp. 1278–1288.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlang. New York, Inc.

Shao, J., Chen, Y., and Chen, Y. (1998). ‘Balanced repeated replication for stratified multistage survey data under imputation,’ *Journal of the American Statistical Associatio*n, **93**, pp. 819–831.

Slootbeek, G. T. (1998). ‘Bias correction in the balanced half-sample method if the number of sampled units in some strata is odd,’ *Journal of Official Statistics*, **14**, pp. 181–188.

Smith, P, Lundström, S. and Underwood, C. (1998b) ‘Evaluation of variance estimation software,’ *Model quality report in business statistics, Vol II: Comparison of Variance Estimation Software and Methods* eds. Davies, P. and Smith, P.

Smith, P., Full, S., Underwood, C., Chambers, R. and Holmes, D. (1998a). ‘Simulation study of alternative variance estimation methods’, in *Model Quality Report in Business Statistics, Vol II: Comparison of Variance Estimation Software and Methods* eds. Davies, P. and Smith, P.

Steel, P. (1997). Annual and quarterly Labour Force Surveys: Report on sample design and estimation procedures. Irish Central Statistics Office.

‘Techniques of variance estimation: the random group method and the jackknife method’ (Technical report, Supcom 1998 project, lot 16)

Valliant, R. (1990). ‘Comparisons of Variance estimators in stratified random and systematic sampling,’ *Journal of Official Statistic*s, **6**, pp. 115–131.

Vehovar, V. (1999). ‘Field substitution and unit nonresponse,’ *Journal of Official Statistics*, **15 (2)**, pp. 335–350.

Wolter, K. (1985). *Introduction to Variance Estimatio*n. Springer, New York.

Yuan, Y. (2000). ‘Multiple imputation for missing values: concepts and new developments.’ SUGI Proceedings, SAS Institute Inc.

Yung, W. and Rao, J. N. K. (2000). ‘Jackknife variance estimation under imputation for estimators using post-stratification information,’ *Journal of the American Statistical Association*, **95**, pp. 903–915.

8. APPENDIX

8.1 Notation

The notation adopted in this report is as follows:

- We have a population of size N, split into H strata
- Stratum h (h=1,..H) is consisted of N_h sampling units, $N = \sum_{h=1}^{H} N_h$
- From stratum h, a sample of size s_h is taken
- Sample s_h is divided into a subset r_h of n_h respondents and a subset nr_h of $(s_h - n_h)$ non-respondents
- Y is the variable of interest with values denoted as y_{hi} (the observation of enterprise i in stratum h)
- The quantities of main interest are totals, means, percentages and ratios and are denoted by *t*, *μ*, *p* and *r*, respectively. If necessary (such as in the case of ratios which is consisted of the totals of two variables), the variable to which they refer appears as superscript (e.g. t^y)
- In the case of GREG weighting:

The estimator of, e.g., a total is defined as $\hat{t}_{GREG} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{i=1}^{n_h} g_{hi} \cdot y_{hi}$

where g_{hi} is the GREG weight defined as $g_{hi} = 1 + x_{hi}^T \hat{A}^{-1} \left(t^x - \hat{t}^x \right)$

> x_{hi} is the vector of the values of the auxiliary variables used in the weighting for the *i* case in stratum *h* (+1 in the first row),
>
> t^x is the (known) vector of totals of the x_{hi} across the population
>
> $\hat{t}^x = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{i=1}^{n_h} x_{hi}$ is essentially the estimation of t^x based on the sample, and
>
> $\hat{A} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{i=1}^{n_h} x_{hi} \cdot x_{hi}^T$

- Apart from the $(s_h - n_h)$ non-respondents (unit non-response), item non-response exists for specific questions and imputation procedures could be applied. The non-imputation rate for the variable *y* is m_y / n_y, where m_y is the number of respondents in the corresponding question of variable *y* and n_y is the number of respondents in the survey that were eligible for answering the specific question.

For the notation of basic concepts one may also refer to any standard textbook, such as Cochran (1977), while GREG weighting is treated in detail in Andersson and Nordberg (1998).