DACSEIS
IST-2000-26057

Workpackage 8

Variance Estimation for estimators subject to Raking Adjustment

Deliverable 8.1

2003

List of contributors:

Julia D'Arrigo and Chris Skinner, Southampton.

Main responsibility:

Julia D'Arrigo and Chris Skinner, Southampton.

IST–2000–26057–DACSEIS

The DACSEIS research project is financially supported within the IST programme of the European Commission. Research activities take place in close collaboration with Eurostat.

http://europa.eu.int/comm/eurostat/research/

http://www.cordis.lu/ist/

http://www.dacseis.de/

Preface

Raking ratio estimation is often used with household surveys in government statistics as a method of calibration and to adjust for nonresponse. This document discusses methods of variance estimation for raking ratio estimators. Two types of raking ratio estimators are considered, together with the generalized regression estimator as a benchmark for comparison. The main focus is on linearization methods, for which the variance estimator for a raking ratio estimator is similar to that for the generalized regression estimator, subject to come choices of weights. Jackknife variance estimation is also considered. A simulation study of the properties of the alternative point estimators and variance estimators is undertaken using data from both the British Labour Force Survey and the German Sample Survey of Income and Expenditure, allowing not only for sampling but also non-response.

Julia D´Arrigo and Chris Skinner Southampton, October 2002

Contents

List of Tables

Chapter 1

Introduction

Raking ratio estimation is one of a number of forms of calibration estimation (Deville *et al.*, 1993). Its main competitor is generalised regression estimation (GREG). Raking ratio estimation appears to have a more well-established history of applications in many national statistical institutes (NSIs), perhaps because of its ease of computation, involving repeated use of standard post-stratification adjustments (Kalton and Flores-Cervantes, 2003). In some NSIs, GREG has tended to replace raking ratio estimation. One reason is that the GREG can be expressed in closed form and computed in one step, whereas the computation of a raking ratio estimator is iterative. Perhaps a more important reason is that GREG can handle a wider class of forms of auxiliary information, including population totals of continuous variables, whereas raking is restricted to the use of population counts in the categories of discrete variables. Nevertheless, raking ratio continues to be used widely in NSIs in many countries, e.g. the USA and the UK. One advantage is that it always produces positive weights, whereas GREG requires modification to meet this condition. In addition, raking may reduce non-response bias more than GREG under certain assumptions (Kalton and Flores-Cervantes, 2003). Although GREG and raking often produce similar estimates and are asymptotically equivalent under certain strong conditions (Deville and Särndal, 1992), their properties still require further comparison, especially in the presence of unit non-response when the conditions of Deville and Särndal (1992) will not hold in general.

This workpackage considers two forms of raking ratio estimation, the classical estimator obtained by the application of iterative proportional fitting as well as an estimator, which may be interpreted as a maximum likelihood estimator within a certain framework (Brackstone and Rao, 1979). The GREG estimator is also considered as a benchmark for comparison. Our primary aim is to investigate alternative variance estimation approaches for the raking ratio estimators. In particular, we focus on linearization variance estimators and consider the choice between using design weights and raking weights both to weight the residuals and to weight the estimated regression coefficients when computing the residuals. We study the properties of these alternative variance estimators both with and without unit non-response. Data from the Great Britain Labour Force Survey (LFS) and the German Sample Survey of Income and Expenditure (SIE), two national surveys, are used to evaluate the properties of these estimators in simulation studies. Both a multiplicative and an additive non-response model are considered in the simulations. The complex designs used for both the LFS and the SIE are mimicked as far as possible in this investigation.

Chapter 2

Raking Ratio Estimators

2.1 Classical Raking Estimator

Consider the estimation of a population total T_Y of a survey variable Y taking values y_i for units i in a population U:

$$T_Y = \sum_U y_i.$$

Given observations on y_i for units in a sample s, a basic weighted estimator of T_Y is given by

$$\widehat{T}_Y = \sum_s \omega_i\, y_i,$$

where ω_i is a given weight, referred to here as the initial weight. This weight may, for example, be the Horvitz-Thompson (H-T) weight $\omega_i = 1/\pi_i$.

The classical raking adjustment makes use of information on the population counts in the categories of two or more categorical auxiliary variables. This type of adjustment is used, for example, in the weighting and calibration procedure of the Britain Labour Force Survey (LFS), where three sets of post-strata are used for calibration. Let x_i denote the vector of indicator variables of these categories, for example in the case of three auxiliary variables:

$$x_i = (\delta_{1..i}, \ldots, \delta_{A..i}, \delta_{.1.i}, \ldots, \delta_{.B.i}, \delta_{..1i}, \ldots, \delta_{..Ci})',$$

where $\delta_{a..i} = 1$ if unit i is in category a of the first auxiliary variable and 0 otherwise, $\delta_{.b.i} = 1$ if unit i is in category b of the second auxiliary variable and 0 otherwise and so on. The population total T_X of this vector thus contains the population counts in each of the (marginal) categories of each of the three auxiliary variables. It is assumed that T_X is given and that x_i is known for $i \in s$.

The classical raking adjustment involves iterative modifications of the initial weights, ω_i, in a multiplicative way to adjusted weights, w_i, with the aim of satisfying the calibration equations:

$$\sum_s w_i x_i = T_X.$$

The resulting raking estimator of T_Y is:

$$\widehat{T}_{Y\,Rak} = \sum_s w_i\, y_i.$$

The multiplicative adjustment, depends only upon the cell in the contingency table formed by the auxiliary variables, that is we may write $w_i = \omega_i h(x_i)$, where the multiplicative adjustment factor $h(x_i)$ is fixed for all units with common values of the auxiliary variables. Let $\widehat{N}_\omega[h(x)]$ and $\widehat{N}_w[h(x)]$ denote the weighted estimates of the population counts in the cell of the table defined by x, using the weights ω_i and w_i respectively. Then we may write $\widehat{N}_w[h(x)] = h(x)\widehat{N}_\omega[h(x)]$. The usual iterative modification of the weights involves iterative proportional fitting (BRACKSTONE and RAO, 1979). IRELAND and KULLBACK (1968) demonstrate that this method converges to a solution which minimises:

$$\sum \widehat{N}_w \log(\widehat{N}_w/\widehat{N}_\omega).$$

subject to the calibration equations, where the sum is over all cells defined by x. This objective function may alternatively be expressed as:

$$\sum_s w_i \log(w_i/\omega_i)$$

that is, under convergence of the iterative algorithm, the w_i minimise the above function, subject to solving the calibration equations. The objective function in this optimisation problem may alternatively be expressed as

$$\sum_s \omega_i G_M(w_i/\omega_i)$$

where $G_M(u) = u\ \log(u) - u + 1$ is the multiplicative distance measure considered by DEVILLE *et al.* (1993), and it is assumed that the calibration equations imply that $\sum w_i$ is constrained to be a given constant. Using the standard Lagrange multiplier method for constrained minimisation, the multiplicative adjustment factors may be expressed as (DEVILLE *et al.*, 1993):

$$w_i = \omega_i F_M(x_i'\widehat{\lambda}),$$

where $F_M(u) = g_M^{-1}(u)$ denotes the inverse function of $g_M(u) = dG_M(u)/du$ and $\widehat{\lambda}$ is the Lagrange multiplier, which solves the calibration equations. It follows from the definition of $G_M(u)$ above that $g_M(u) = \log(u)$ and $F_M(u) = \exp(u)$. Hence $\widehat{\lambda}$ solves

$$\sum_s \omega_i \exp(x_i'\widehat{\lambda})x_i = \sum_U x_i.$$

In the discussion in this chapter, the estimator $\widehat{T}_{Y\,Rak} = \sum w_i\, y_i$ may be used if y_i is a scalar or a vector, since the scalar weights, w_i, do not depend upon y_i. More generally, if θ is a population parameter which may be expressed as a function of a vector of population totals $\sum y_i$ then the corresponding function of the weighted sample sum $\sum w_i y_i$ provides a raked estimator of θ.

2.2 'Maximum Likelihood' Raking Estimator

An alternative approach to raking adjustment involves a maximum likelihood argument for the estimation of the population proportions in the cells of the table formed by cross-classifying the auxiliary variables. The calibration equations remain the same but the objective function, summed over the cells, is replaced by (Brackstone and Rao, 1979; Fuller, 2002):

$$-\sum \widehat{N}_\omega \log(\widehat{N}_w),$$

which is proportional to minus the log likelihood in the case of simple random sampling with replacement (Brackstone and Rao, 1979). Equivalently, the objective function may be expressed, summed over sample units, as:

$$\sum_s -\omega_i \log(w_i/\ \omega_i),$$

(omitting an additional term not involving w_i), or as (Fuller, 2002):

$$\sum_s \omega_i G_{ML}(w_i/\ \omega_i),$$

where $G_{ML}(u) = u - 1 - \log(u)$. The function G_{ML} is Case 4 of the distance measures considered by Deville and Särndal (1992).

2.3 Generalised Regression Estimator

As a benchmark for comparing the properties of the raking estimators, we also consider the widely used generalised regression estimator (GREG).

Deville and Särndal (1992) and Deville *et al.* (1993) discuss other choices G for the functions G_M and G_{ML} considered above. For every fixed ω_i, it is assumed that G is nonnegative, differentiable with respect to w_i, strictly convex in order for a unique solution to exist for the optimisation problem, and $G(1) = 0$ implying that when $w_i = \omega_i$ the distance between the weights is zero. Additionally, it is required that G' is continuous, one-to-one, and that G'$(1) = 0$ and G"$(1) > 0$ which makes $w_i = \omega_i$ a local minimum.

An alternative objective function, which leads to the GREG estimator, is:

$$\sum_s \omega_i G_{LM}(w_i/\ \omega_i),$$

where $G_{LM}(u) = (1/2)(u-1)^2$. The function G_{LM} is Case 1 of the distance measures considered by Deville and Särndal (1992), referred as the linear method, and this leads to the generalised regression estimator

$$\widehat{T}_{YGreg} = \sum_s w_i y_i = \widehat{T}_{Y\pi} + (T_X - \widehat{T}_{X\pi})' \widehat{B}_s,$$

where $\widehat{T}_{Y\pi} = \sum_s (1/pi_i) y_i$ and $\widehat{T}_{X\pi} = \sum_s (1/pi_i) y_i$ denote the Horvitz-Thompson estimators for Y and the X-vector; and

$$\widehat{B}_s = \left(\sum_s \omega_i x_i x_i^{`} \right)^{-1} \sum_s \omega_i x_i y_i$$

is a weighted estimator of the multiple regression coefficient.

A disadvantage of GREG estimation compared to raking estimation is that, as noted by Deville and Särndal (1992), the adjusted weights, w_i, resulting by using the objective function G_{LM} can be positive or negative, whereas classical raking adjustment and 'maximum likelihood' raking adjustment guarantee positive weights.

Chapter 3

Variance Estimation by Linearization

3.1 The Asymptotic Variance

Following BINDER and THÉBERGE (1988) and DEVILLE *et al.* (1993), we consider estimating the asymptotic variance of the 'converged' estimator, i.e. the estimator $\widehat{T}_{YRak}$, where the weights w_i solve the constrained optimisation problem. This asymptotic variance is assumed to be a sufficiently close approximation to the asymptotic variance of the estimator obtained after the finite number of iterations used in practice.

We assume that in large samples, $\widehat{\lambda}$ converges to a value λ. DEVILLE and SÄRNDAL (1992) assume that $\lambda = \mathbf{0}$, but this property is based upon the assumption that the estimator of T_X, obtained by applying the basic weights ω_i, is consistent. This assumption will often be false in the case of non-response and we prefer not to make this assumption.

We allow the function $F(.)$ to be general and not necessarily equal to $F_M(.)$. We first expand the adjusted weight $w_i = \omega_i F(x_i'\widehat{\lambda})$ about λ to obtain

$$w_i \approx \omega_i[F_i + f_i x_i'(\widehat{\lambda} - \lambda)],$$

where $F_i = F(x_i'\lambda)$, $f_i = f(x_i'\lambda)$, $f(u) = dF(u)/du$. Substituting in the calibration equations we obtain:

$$\sum_s \omega_i[F_i + f_i x_i'(\widehat{\lambda} - \lambda)]x_i \approx T_X,$$

and hence

$$\widehat{\lambda} - \lambda \approx \left[\sum_s \omega_i f_i x_i x_i'\right]^{-1} \left[T_X - \sum_s \omega_i F_i x_i\right].$$

We are assuming here that the first matrix in this expression is non-singular. It may be necessary to drop redundant variables from x_i to achieve this. For example, in the three-way case above, each of the sums of the indicator variables $\delta_{a..i}$, $\delta_{.b.i}$ and $\delta_{..ci}$ across a, b and c, respectively, equals 1 and it is natural to drop two of these indicators to avoid singularity.

Substituting in the estimator of interest we obtain:

$$\widehat{T}_{YRak} \approx \sum_s \omega_i \left[F_i + f_i x_i^{`}(\widehat{\lambda} - \lambda)\right] y_i \approx \sum_s \ddot{w}_i y_i + B\left[T_X - \sum_s \ddot{w}_i x_i\right],$$

where $\ddot{w}_i = \omega_i F_i$ and $B = \left[\sum_s \omega_i f_i y_i x_i^{`}\right]\left[\sum_s \omega_i f_i x_i x_i^{`}\right]^{-1}$.

We assume that B converges to a finite limit matrix β in the asymptotic framework. It follows from this and the other approximating assumptions above, in particular that the basic weights ω_i are fixed, that the (normalized) asymptotic distribution of $B[T_X - \sum_s \ddot{w}_i x_i]$ is the same as that of $\beta[T_X - \sum_s \ddot{w}_i x_i]$. Hence the (normalized) asymptotic variance of $\widehat{T}_{YRak}$ is the same as that of $\sum_s z_i$, where z_i is the linearized variate

$$z_i = \ddot{w}_i\,(y_i - \beta x_i)$$

3.2 Linearization Variance Estimator

For the purpose of linearization variance estimation, $\widehat{T}_{YRak}$ may be treated as the linear estimator $\sum_s \widehat{z}_i$, where $\widehat{z}_i = w_i(y_i - \widehat{B}x_i)$ is treated as a fixed variable and $\widehat{B}$ is defined in the same way as B, with f_i replaced by $\widehat{f}_i = f(x_i^{'}, \widehat{\lambda})$. The variance estimator for $\widehat{T}_{YRak}$ is then obtained for a given sampling scheme by using a standard variance estimator for that scheme for a linear estimator, applied to $\sum_s \widehat{z}_i$. The same approach applies whether y_i is a scalar or a vector. For an estimator $\widehat{\theta}$ which is a smooth function of a weighted sample sum, the estimator may first be linearized by the usual linearization techniques to create a linearized variate u_i for which the variance of $\widehat{\theta}$ may be approximated by the variance of $\sum w_i u_i$, treating the w_i as fixed quantities. To allow for raking adjustment in the weights w_i, the standard variance estimator for the linear estimator $\sum_s \widehat{z}_i$ may again be applied, with u_i replacing y_i in the definition of $\widehat{z}_i$.

DEVILLE and SÄRNDAL (1992) note that in their classical theory the choices $\widehat{z}_i = \omega_i(y_i - \widehat{B}x_i)$ or $\widehat{z}_i = w_i(y_i - \widehat{B}x_i)$ are asymptotically equivalent but they express a preference for the second choice.

The definition of $\widehat{B}$ above in the formula for $\widehat{z}_i$ gives the same variance estimator proposed by DEMNATI and RAO (2004).

For example, for a stratified sampling design assuming "with replacement" sampling of clusters within strata, as in the LFS design, a standard estimator of the variance of $\widehat{T}_{YRak}$ is given by:

$$\widehat{V}(\widehat{T}_{Y\,Rak}) = \sum_{h=1}^{H} \frac{n_h}{n_h - 1} \sum_{j=1}^{n_h} (z_{hj} - \overline{z}_h)^2 \qquad (3.1)$$

where $z_{hj} = \sum_k z_{hjk}$, $\overline{z}_h = \sum_j z_{hj}/n_h$, $z_{hjk} = w_{hjk}y_{hjk}$, and z_{hjk} is the value of the variable z for the kth individual within the jth selected cluster in stratum h. This standard variance estimator is based on the assumption that w_{hjk} are fixed. To allow for raking adjustment in the weights w_{hjk}, the *standard linearization* variance estimator of $\widehat{T}_{YRak}$ is given by (3.1) with z_{hj} defined as:

$$z_{hj} = \sum\nolimits_k \omega_{hjk}\, e_{hjk} \tag{3.2}$$

where $e_{hjk} = y_{hjk} - x'_{hjk}, \widehat{B}_s$ are the estimated residuals and $\widehat{B}_s$ is the estimator of the multiple regression coefficient, which may be constructed using either design (Deville and Särndal, 1992) or raking weights. An alternative variance estimation method for $\widehat{T}_{YRak}$ is the *jackknife linearization* variance estimator. That is obtained by using (3.1) with z_{hj} defined as:

$$z_{hj} = \sum\nolimits_k w_{hjk}\, e_{hjk} \tag{3.3}$$

In the case of classical raking adjustment, $F(u) = F_M(u) = \exp(u)$. Hence, $f(u) = f_M(u) = F_M(u)$ and $w_i = \omega_i \widehat{f}_i$, and so $\widehat{B} = \widehat{B}_s$ and the two variance estimators are identical. For the 'maximum likelihood' choice $G_{ML}(u) = u - 1 - \log(u)$, we have $F_{ML}(u) = (1-u)^{-1}$ and $f_{ML}(u) = (1-u)^{-2}$ so that $\omega_i \widehat{f}_i = w_i^* = w_i^2/\omega_i$ and the two variance estimators are not identical.

Chapter 4

Jackknife and other Replication Variance Estimators

Let $\widehat{\theta}$ be an estimator of a scalar population parameter θ, based upon a set of raked weights w_i. The raking method may be any of those considered in Chapter 2. The estimator of the variance of $\widehat{\theta}$ for a broad class of replication methods is obtained by first constructing a set of T replicates weights $w_i^{(t)}$ for $t = 1, \ldots, T$, according to the replication method (Rust and Rao, 1996). Methods of construction are discussed below. For each set of replicate weights, an estimator $\widehat{\theta}^{(t)}$ of θ is then computed in the same way that $\widehat{\theta}$ is computed using the weights w_i. An estimator of the variance of $\widehat{\theta}$ is then given by

$$\widehat{V}(\widehat{\theta}) = \sum_{t=1}^{T} c_t (\widehat{\theta}^{(t)} - \widehat{\theta})^2 \tag{4.1}$$

where c_t is a constant which depends on the replication method. For the bootstrap method $c_t = 1/T$. For the jackknife method, c_t is defined below.

The construction of the replicate weights $w_i^{(t)}$ involves first taking the initial weights ω_i. From these a set of initial replication weights $\omega_i^{(t)}$, $t = 1, \ldots, T$, is constructed according to the replication method and the sampling scheme. Next the raking adjustment method is applied to each of these T sets of weights separately. This generates the required weights $w_i^{(t)}$. This approach can be applied validly to a wide class of adjustment methods including classical raking, the "maximum likelihood" raking and the generalized regression estimation (see Rust and Rao, 1996).

To illustrate the construction of the initial replicate weights $\omega_i^{(t)}$, consider the jackknife method with stratified multistage sampling. The number of replicates in this case is $T = \sum n_h$, where n_h is the number of primary units (PSUs) in stratum $h = 1, \ldots, H$ and H is the number of strata. Let replicate t correspond to deleting PSU j in stratum h and let the label i for the weight ω_i correspond to element m in PSU l in stratum k. Then the replicate weights are given by

$$\omega_i^{(t)} = \begin{cases} \omega_i & \text{if } k \neq h \\ \omega_i / c_t & \text{if } k = h, l \neq j \\ 0 & \text{if } k = h, l = j \end{cases} \tag{4.2}$$

where $c_t = (n_h - 1)/n_h$, for $t = 1, \ldots, T$.

The jackknife method described above requires that each cluster within each stratum is deleted in turn. This could require many recalculations for large surveys and thus be prohibitive. An alternative is to group the n_h clusters in the hth stratum into $g_h \geqslant 2$ groups ($g_h < n_h$) and to proceed as if these were the actual clusters. Thus each group is deleted in turn and the number of recalculations is reduced to $\sum_h g_h$. The only change in (4.1) is in c_t as follows:

$$c_t = \frac{g_h - 1}{g_h} \tag{4.3}$$

Both of the above methods require the raking method to be applied to each set of weights. In principle, this requires iterating the raking method until convergence in each case. This imposes a high computational burden. An alternative is to reduce the number of iterations of the raking method, in particular by using a one-step jackknife (SHAO and TU, 1995, p.191). One version of the one-step jackknife, adopted by CANTY and DAVISON (1999), is simply to stop after 'one step' of the raking adjustment to the initial replication weights, rather than continuing to convergence. This one step might consist of one step of Newton's method (DEVILLE *et al*, 1993). CANTY and DAVISON (1999) compare the performance of their one-step jackknife method with the jackknife method involving five iterations. They find that the performance of the variance estimator is actually worse for five iterations and conclude that "overall, the best jackknife strategy appear to be to use one iteration" (p.387). Alternatively, the one step of the raking adjustment could be applied to a set of weights $w_i^{(t)}$ formed by replacing the initial weights ω_i in (4.2) by the raked weights w_i. These approaches should be asymptotically equivalent (SHAO and TU, 1995) but their finite sample properties require further study. Either of these one-step methods still requires the inversion of a matrix in the one step of Newton's method for each replicate and this could still be computationally heavy. This repeated inversion of a matrix could be avoided by use of the estimation function jackknife, studied by RAO and TAUSI (2003).

Chapter 5

Simulation Study

In order to compare the performance of the calibration estimators presented in Chapter 2 and their corresponding variance estimators, discussed in Chapter 3 and 4, we undertake two simulation studies, using data from the British Labour Force Survey and the German Sample Survey of Income and Expenditure, in which we investigate their finite sample design-based properties. See workpackage 2 for further details.

5.1 The British Labour Force Survey

The British LFS is a quarterly survey of persons living in private households in Britain. Its purpose is to provide information on the British labour market which can then be used to develop, manage, evaluate and report on labour market policies. It is carried out by the Social Survey Division of the Office for National Statistics (ONS).

5.1.1 Population and Sampling Design

Data from the basic quarterly LFS in Great Britain was used as our artificial population. The LFS is a very large survey which results in approximately 58,000 addresses in our artificial population. In particular, we consider the March-May 1998 LFS sample data. From this data repeated samples were drawn in a way which mimics as far as possible the design used for the LFS. The details of the design of the survey can be found in Office for National Statistics (1998), Section 3.

The sample consists of 1,211 simple random sampling addresses (= clusters of individuals) allocated proportional into 19 strata, defined by region of residence, giving on average a sampling fraction of about 1/48 for our artificial scheme. The distribution of the population and sample sizes across the strata, and the sampling fraction can be found in Appendix A.

The regions of residence are used to mimic the effect of the 110 Interviewer Areas (IAs) used in the LFS itself. The IAs of the LFS can be considered homogeneous strata and so we conceive our region divisions as equivalent homogeneous strata for our samples. A probability simple random sample (SRS) is selected within each stratum.

In the LFS all individuals in a sampled address are interviewed if possible. In this simulation study, all the respondents in a sample address are retained, except those aged under 16, who are not relevant for the estimates of interest.

Two different non-response models were taken into account in this study, a multiplicative and an additive. Information to assign non-response probabilities, ϕ_{hj}, to each selected cluster of the population is obtained from FOSTER (1998), and takes into account characteristics of clusters, such as area of residence, age of head of household and sex of head of household. Once the sample is drawn, in order to obtain the subset of respondents, Bernoulli distributions with parameter ϕ_{hj}, for $h = 1, \ldots, H$ and $j = 1, \ldots, J_h$, are used to generate an indicator variable I_{hj} that takes value 1 if cluster (hj) responds and value 0 is cluster (hj) does not respond.

Multiplicative Non-response Model:

Pr(non-response) = ϕ_{hj} = 0.12 * 2 if London *1.5 if under 35 * 1.4 if female

Additive Non-response Model:

Pr(non-response) = ϕ_{hj} = 0.12 + 0.12 if London + 0.09 if under 35 + 0.06 if female

5.1.2 Weighting and Calibration

Each sampled individual, within an address, is assigned a weight. To account for non-response and noncoverage, these weights are adjusted by using auxiliary information, available from an external source. The weights are constructed by raking to three sets of population control totals that include region, age and sex. Classical raking adjustment is used via iterative poststratification to each set of control totals in turn. We try to mimic this structure, as far as possible, in our study. However, because of the problem of strata with small numbers of individuals that occurs due to our artificial population and samples are much smaller than those for the LFS, we modify the LFS post-strata as follows:

- The first set of post-strata is the area of residence with 23 levels, which are given in Table 5.1.

- The second set of post-strata is a cross-classification of sex by age, in single years, for those between 16 and 24 and 25 or older. This gives us a factor with 20 levels.

- The third set of post-strata is a cross-classification of region (1. Northern England, 2. London and South East, 3. The Midlands and East Anglia and 4. Scotland) by sex by age, in 15-year age groups, 16-29, 30-44, 45-59, 60-75 and 75 or older. This gives us a factor with 40 levels.

Table 5.1: Areas of Residence for post-stratification

Area	*Counties*
1	Cleveland, Cumbria and Durham
2	Northumberland, Tyne & Wear and Humberside
3	North Yorkshire, West Yorkshire and South Yorkshire
4	Derbyshire, Leicestershire and Nottinghamshire
5	Lincolnshire, Northamptonshire and Cambridgeshire
6	Norfolk and Suffolk
7	Bedfordshire, Hertfordshire and Essex
8	Inner London and Outer London
9	East Sussex, West Sussex, Kent and Surrey
10	Hampshire, Isle of Wight, Dorset and Wiltshire
11	Berkshire, Buckinghamshire and Oxfordshire
12	Avon, Gloucestershire and Somerset
13	Cornwall and Devon
14	Hereford & Worcester, Shropshire and Staffordshire
15	Warwickshire and West Midlands
16	Cheshire and Merseyside
17	Greater Manchester and Lancashire
18	Clwyd, Dyfed, Gwent and Gwynedd
19	Mid Glamorgan, South Glamorgan, West Glamorgan and Powys
20	Border, central and Dumfries & Galloway
21	Fife and Grampian
22	Highland, Tayside and Northern & Western Isles
23	Lothian and Strathclyde

5.1.3 Statistics

The parameters of interest are the total number of unemployed (TNU), the total number of employed (TNE) and the total number in the inactive workforce (TNI). These parameters are computed from the finite artificial population of the LFS by $T_y = \sum_{(hjk)\in U} y_{hjk}$, where $\boldsymbol{y}_{hjk}$ is the vector of indicator survey variables $\boldsymbol{y}_{hjk} = (y_{1hjk}, y_{2hjk}, y_{3hjk})$, where $y_{1hjk} = 1$ if individual (hjk) is unemployed and 0 otherwise, $y_{2hjk} = 1$ if individual (hjk) is employed and 0 otherwise, and $y_{3hjk} = 1$ if individual (hjk) is inactive and 0 otherwise. For each of the $R = 1,000$ samples, estimates of the total unemployed, employed and inactive were calculated using the classical raking estimator, the "ML" raking estimator, and the GREG estimator, with $\boldsymbol{y}_{hjk}$ defined as above. Two models of non-response as well as complete response are considered in this study. The calibration weights w_{hjk} were determined by using the distance functions G_M, G_{ML}, and G_{LM} presented in Chapter 2. That implies a total of twenty seven point estimators.

For each of the $R = 1,000$ samples and each of the calibration point estimators, we compute the standard linearization variance estimator given by (3.1) and (3.2), the jackknife linearization variance estimator given by (3.1) and (3.3), and the grouped jackknife replication variance estimator given by (4.1) and (4.2).

A number of statistics and properties are investigated in this simulation study as follows:

(1) The *expectation of the point estimator* $\widehat{T}_y$ (TNU, TNE or TNI) is estimated by:

$$\widehat{\mathrm{E}}(\widehat{T}_y) = \frac{1}{R}\sum_{r=1}^{R} \widehat{T}_{y_r},$$

where $\widehat{T}_{y_r}$ is the value of $\widehat{T}_y$ for sample r.

(2) The ("true") *variance of the point estimator* $\widehat{T}_y$ is estimated by:

$$V_{\mathrm{TRUE}} = \frac{1}{R}\sum_{r=1}^{R}\left[\widehat{T}_{y_r} - \widehat{\mathrm{E}}(\widehat{T}_y)\right]^2.$$

(3) The *percent relative bias of the point estimator* $\widehat{T}_y$, relative to the population value T_y, is estimated by:

$$RBias(\widehat{T}_y) = \frac{\frac{1}{R}\sum\limits_{r=1}^{R}(\widehat{T}_{y_r} - T_y)}{T_y} * 100 = \frac{Bias\,(\widehat{T}_y)}{T_y} * 100.$$

(4) The *variance of the bias estimator* is estimated by:

$$V\left[Bias\,(\widehat{T}_y)\right] = \frac{V_{\mathrm{TRUE}}}{R}.$$

(5) The *root mean square error of the point estimator* $\widehat{T}_y$ is estimated by:

$$RMSE(\widehat{T}_y) = \sqrt{V_{\mathrm{TRUE}} + \left[Bias\left(\widehat{T}_y\right)\right]^2} = \sqrt{MSE(\widehat{T}_y)}.$$

(6) The *expectation of a variance estimator* $\widehat{V}(\widehat{T}_y)$ for $\widehat{T}_y$ is estimated by:

$$\widehat{\mathrm{E}}\left[\widehat{V}(\widehat{T}_y)\right] = \frac{1}{R}\sum_{r=1}^{R}\widehat{V}_r(\widehat{T}_y),$$

where $\widehat{V}_r(\widehat{T}_y)$ is the value of $\widehat{V}(\widehat{T}_y)$ for sample r.

(7) The *percent relative bias of a variance estimator* $\widehat{V}(\widehat{T}_y)$, relative to the "true" variance, is estimated by:

$$RBias\left[\widehat{V}(\widehat{T}_y)\right] = \frac{\frac{1}{R}\sum\limits_{r=1}^{R}\left[\widehat{V}_r(\widehat{T}_y) - V_{\mathrm{TRUE}}\right]}{V_{\mathrm{TRUE}}} * 100 = \frac{Bias\left[\widehat{V}(\widehat{T}_y)\right]}{V_{\mathrm{TRUE}}} * 100.$$

(8) The *variance of a variance estimator* $\widehat{V}(\widehat{T}_y)$ is estimated by:

$$V\left[\widehat{V}(\widehat{T}_y)\right] = \frac{1}{R}\sum_{r=1}^{R}\left\{\widehat{V}_r(\widehat{T}_y) - \mathrm{E}\left[\widehat{V}(\widehat{T}_y)\right]\right\}^2.$$

(9) The *variance of the estimated bias of a variance estimator* $\widehat{V}(\widehat{T}_y)$ is estimated by:

$$V\left\{Bias\left[\widehat{V}(\widehat{T}_y)\right]\right\} = \frac{V\left[\widehat{V}(\widehat{T}_y)\right]}{R}.$$

(10) The *root mean square error of a variance estimator* $\widehat{V}(\widehat{T}_y)$ is estimated by:

$$RMSE\left[\widehat{V}(\widehat{T}_y)\right] = \sqrt{V\left[\widehat{V}(\widehat{T}_y)\right] + \left\{Bias\left[\widehat{V}(\widehat{T}_y)\right]\right\}^2} = \sqrt{MSE\left[\widehat{V}(\widehat{T}_y)\right]}.$$

(11) The *percent coefficient of variation of a variance estimator* $\widehat{V}(\widehat{T}_y)$ is estimated by:

$$CV\left[\widehat{V}(\widehat{T}_y)\right] = \frac{\frac{1}{R}\sum_{r=1}^{R}\left\{\widehat{V}_r(\widehat{T}_y) - \mathrm{E}\left[\widehat{V}(\widehat{T}_y)\right]\right\}^2}{V_{\mathrm{TRUE}}} * 100.$$

(12) The *coverage of the confidence interval* $\widehat{T}_{y_r} \pm 1.96\sqrt{\widehat{V}_r(\widehat{T}_y)}$, constructed using the point estimator $\widehat{T}_{y_r}$, the variance estimator $\widehat{V}_r(\widehat{T}_y)$ (standard or jackknife) and a nominal 95% level, is estimated by the percentage of these R intervals ($r = 1, \ldots, R$) which contain the true value T_y. The desired value of this coverage is 95%.

Finally, some statistics related to the calibration weights are computed for each of the distance function under study, such as the number of negative weights, the number of calibration weights less than one, the number of calibration weights more than 10 times the design weights, the mean of the coefficient of variation of the calibration weights and the mean of the chi-square distance between the calibration and the design weights.

5.2 The German Sample Survey of Income and Expenditure

The sample survey of income and expenditure of household is a nationwide survey conducted every 5 years by the Federal Statistical Office. The main purpose of the survey is to represent the economic and social situation of households with respect to income distribution and use. For further information see workpackage 2 and `http://www.destatis.de/download/veroe/methoden.pdf`.

5.2.1 Population and Sampling Design

Data from the German SIE, in particular from the federal state of Bremen, was used to produce an artificial population. According to the Microcensus 1995, Bremen has a population of 344,600 households. For computational reasons, such as lack of memory, we consider a simple random sample of 68,714 (20% sampling fraction) households from Bremen as our artificial population. As households with a monthly household net income of DM 35,000 and over do not belong to the German SIE population, they were excluded from the simulated population resulting in a final artificial population of 64,326 households. From this data repeated probability simple random samples (SRS) of 1,340 households were drawn, giving a sampling fraction of about 1/48 for our artificial scheme.

Even though the SIE sampling design, combined quota sampling, does not allow for non-response, two different non-response models were taken into account in this study for research purposes, a multiplicative and an additive. Information to assign non-response probabilities, ϕ_i, to each selected household of the population is obtained from the Family Expenditure Survey and the National Food Survey of Great Britain (FOSTER, 1998). This information takes into account characteristics of households, such as socio-economic status and type of household. Once the sample is drawn, to obtain the subset of respondents, Bernoulli distributions with parameter ϕ_i, for $i = 1, \ldots, I$, are used to generate an indicator variable I_i that takes value 1 if household i responds and value 0 if household i does not respond.

Multiplicative Model:

Pr(non-response) = ϕ_i = 0.304 * 1.80 if self-employed * 1.06 if unemployed * 0.94 if employed * 1.30 if no children in the household

Additive Model:

Pr(non-response) = ϕ_i = 0.304 + 0.055 if self-employed + 0.017 if unemployed – 0.017 if employed + 0.096 if no children in the household

5.2.2 Weighting and Calibration

As for the LFS, each sampled household is assigned a weight, which is computed considering auxiliary information from an external source. In the SIE the weights are constructed using essentially the maximum likelihood raking method by adjusting the sample data simultaneously to the marginal distributions of several characteristics, such as household type, social economic status of the reference person, household net income class and land. We try to mimic this adjustment, as far as possible, in our study. However, as for the LFS, because of the problem of strata with small numbers of households we modify the SIE calibration variables as follows:

- The first set of post-strata is the household type with 7 levels, as follows:

0 ...	Other
1 ...	mother/father alone + 1 child
2 ...	mother/father alone + 2 or more children
3 ...	couple with 1 child – spouse employed
4 ...	couple with 1 child – spouse unemployed
5 ...	couple with 2 or more children – spouse employed
6 ...	couple with 2 or more children – spouse unemployed

- The second set of post-strata is the social status of the reference person with 5 levels, as follows:

0 ...	self employed
1 ...	civil servant or military
2 ...	Employee
3 ...	Worker
4 ...	unemployed, pensioner, student, others

- The third set of post-strata is the household net income class with 3 levels, as follows:

0 ...	from 0 to under 5,000 DM
1 ...	from 5,000 to under 7,000 DM
2 ...	from 7,000 to under 35,000 DM

5.2.3 Statistics

The parameters of interest are the total household net income (INC) and the total household expenditure (EXP). These parameters are computed from the artificial finite population of the SIE by $T_Y = \sum_{i \in U} y_i$, where y_i is the value of the continuous survey variable (INC or EXP) for household i. For each of the $R = 1,000$ samples, the total estimates of income and expenditure were calculated by using the two raking ratio estimators presented in Chapter 2 and the generalised regression estimator. As for the LFS, two models of non-response as well as complete response are considered in this study. The calibration weights, w_i, were determined by the three different distance functions under study, G_M, G_{ML} and G_{LM}.

For each of the $R = 1,000$ samples and each of the calibration point estimators, we compute the standard linearization variance estimator, the jackknife linearization variance estimator, and the grouped Jackknife variance estimator (Chapter 3 and 4).

A number of statistics and properties are investigated in this simulation study, such as relative bias, standard deviation, root mean square error, percent coefficient of variation and confidence interval coverage (See "Statistics" for LFS).

Chapter 6

Results

6.1 The British Labour Force Survey

6.1.1 Properties of the alternative point estimators under different non-response models

Table 6.1 presents biases, standard errors, root mean square errors, numbers of negative weights and very large weights (defined as calibration weights at least 10 times larger than the design weight) for alternative estimators of the total number of unemployed people in the UK, using the LFS data and considering the two non-response models presented in Chapter 5, Subsection 5.1.1, as well as complete response. Numbers are rounded to the nearest two decimals.

One observation about this table is that the standard error remains virtually constant across the estimators within each non-response model. Allowing for non-response, an increase in the simulation standard error across all estimators is observed. Regarding bias, the table shows that for all the point estimates the bias is not significant ($z = \mid \text{bias} \mid / \text{se(bias)} < 1.96$), even under the non-response models, and the root mean square error is very similar to the standard error in each case.

Under non-response, the G_{LM} function, which leads to the GREG estimator, generate some negative adjusted weights while the raking functions, G_M and G_{ML}, seem to guarantee positive weights. Some very large weights are observed for the 'maximum likelihood' raking estimator.

6.1.2 Properties of the variance estimators for the alternative calibration methods under different non-response models

The average value of each standard error estimator $\widehat{S}\text{E} = \sqrt{\widehat{V}}$ for the alternatives total number of unemployed people estimators, under different non-response models and complete response, is shown in the Table 6.2 This table also summarizes the estimation bias for each standard error estimator, the root mean square error of these standard errors

Table 6.1: Simulation Bias of Estimators of Total Unemployed with associated Standard Error, Root Mean Square Error, number of negative and very large calibration weights, R = 1,000, data from LFS.

Non-response Model/ Point Estimator	Bias (Standard error)	"True" Standard Error	Root Mean Square Error	Number of Negative Weights	Number of Very Large Weights
Complete Response:					
GREG	-3.29 (14.68)	464.11	464.12	0	0
Classical Raking	-3.29 (14.70)	464.74	464.75	0	0
'ML' Raking	-3.54 (14.73)	465.74	465.76	0	0
Multiplicative non-response:					
GREG	-23.04 (16.11)	509.38	509.90	11	0
Classical Raking	-23.08 (16.13)	510.04	510.56	0	0
'ML' Raking	-22.81 (16.17)	511.24	511.75	0	7
Additive non-response:					
GREG	-26.51 (16.14)	510.24	510.93	14	0
Classical Raking	-26.53 (16.16)	511.13	511.82	0	0
'ML' Raking	-26.39 (16.21)	512.63	513.31	0	6

and the coverage confidence interval percentage. Since the grouped jackknife method is computationally intensive and extremely time consuming, we only used this method for the case of classical raking estimation, complete response and multiplicative non-response.

Three facts are at once obvious: (1) for the complete response case, the studied estimators are similar to each other and the simulation standard error (see Table 6.1), for each of the considered variance estimation method, with estimation biases relative small compared to the simulation standard error. The jackknife linearization method every time has smaller bias and RMSE of standard error than the standard linearization method. The grouped jackknife method, when applicable, results in significantly smaller bias estimation but higher RMSE than those for the other methods; (2) for the non-response cases, the standard linearization method considerable underestimates the standard error. An important reduction in the bias an RMSE of the standard error estimator is observed when the jackknife linearization method is applied instead of the standard linearization method. Once again, the grouped jackknife standard error estimator, for the raking estimator, shows a considerable smaller estimation bias but the highest RMSE. Because of the occurrence of negative adjusted weights for the GREG estimator under non-response, the standard and jackknife linearization methods can not be applied; (3) little change is observed in the standard error estimators when calibrated, $w_i or w_i^*$, instead of initial weights, ω_i, are used to compute the estimated regression coefficients; however, no clear pattern is observed.

The last column of Table 6.2 gives the empirical coverage of the confidence intervals. For the complete response case, the coverage levels are similar for all the linearization method estimators, ranging in values from 92.8% to 93.8%, and slightly higher for the grouped jackknife standard error estimator (94.70). The coverage levels corresponding to the jackknife linearization method are always slightly larger than that of standard linearization method. Under non-response, confidence levels when using linearization

Table 6.2: Expected Standard Error Estimate for Total Unemployed with associated Bias, Root Mean Square Error and Coverage, R = 1,000, data from LFS.

Estimator	Variance Estimation Method	Expected Standard Error Estimate	Bias of SE Estimate (s.e.)	RMSE of SE Estimate	Coverage of Confidence Interval (%)
Complete Response:					
GREG	Std. Lin. (1)	433.77	-30.34 (0.90)	41.67	92.90
	Std. Lin. (2)	434.07	-30.04 (0.90)	41.47	92.90
	Jackk. Lin. (1)	441.38	-22.73(0.98)	38.40	93.60
	Jackk. Lin. (2)	440.51	-23.59 (0.97)	38.79	93.60
Classical Raking	Std. Lin. (1)	433.77	-30.97 (0.90)	42.13	92.80
	Std. Lin. (2)	434.06	-30.68 (0.90)	41.93	92.80
	Jackk.Lin. (1)	441.48	-23.25 (0.98)	38.83	93.60
	Jackk. Lin. (2)	440.52	-24.22 (0.98)	39.26	93.60
	Grouped Jackk.	461.69	-5.67(1.39)	44.24	94.70
'ML' Raking	Std. Lin. (1)	433.77	-31.98 (0.90)	42.88	93.10
	Std. Lin. (2)	434.07	-31.67 (0.90)	42.66	93.20
	Std. Lin. (3)	435.13	-30.61 (0.91)	41.93	93.40
	Jackk. Lin. (1)	441.97	-23.77 (1.00)	39.46	93.80
	Jackk. Lin. (2)	440.66	-25.08 (0.98)	39.97	93.70
	Jackk. Lin. (3)	440.04	-25.70 (0.98)	40.20	93.70
Multiplicative non-response:					
GREG	Std. Lin. (1)	391.88	-117.50 (0.91)	120.95	87.00
	Std. Lin. (2)	–	–	–	–
	Jackk. Lin. (1)	484.74	-24.65 (1.19)	45.09	93.50
	Jackk. Lin. (2)	–	–	–	–
Classical Raking	Std. Lin. (1)	391.88	-118.16 (0.91)	121.59	87.10
	Std. Lin. (2)	392.30	-117.74 (0.91)	121.20	87.10
	Jackk. Lin. (1)	485.06	-24.98 (1.20)	45.54	93.10
	Jackk. Lin. (2)	483.19	-26.85 (1.19)	46.26	92.90
	Grouped Jackk.	519.05	9.53(7.50)	198.55	94.14
'ML' Raking	Std. Lin. (1)	391.88	-119.35 (0.91)	122.76	87.30
	Std. Lin. (2)	392.33	-118.91 (0.91)	122.32	87.50
	Std. Lin. (3)	393.98	-117.25 (0.91)	120.73	87.70
	Jackk. Lin. (1)	486.39	-24.85 (1.24)	46.37	93.00
	Jackk. Lin. (2)	483.48	-27.76 (1.20)	47.14	93.00
	Jackk. Lin. (3)	482.13	-29.11 (1.19)	47.57	93.00

Table 6.2 continued on next page

methods ranging in values from 86.80% to 93.50% are observed. The coverage level when the grouped jackknife is applied is 94.14%. All the time, the standard method results in narrower confidence intervals than the jackknife methods.

Continuation of Table 6.2

Estimator	Variance Estimation Method	Expected Standard Error Estimate	Bias of SE Estimate (s.e.)	RMSE of SE Estimate	Coverage of Confidence Interval (%)
Additive non-response:					
GREG	Std. Lin. (1)	390.64	-119.60 (0.91)	122.99	87.30
	Std. Lin. (2)	–	–	–	–
	Jackk. Lin. (1)	485.38	-24.86 (1.20)	45.38	93.10
	Jackk. Lin. (2)	–	–	–	–
Classical Raking	Std. Lin. (1)	390.64	-120.49 (0.91)	123.85	87.10
	Std. Lin. (2)	391.06	-120.07 (0.91)	123.46	87.10
	Jackk. Lin. (1)	485.69	-25.44 (1.21)	45.98	92.80
	Jackk. Lin. (2)	483.84	-27.29 (1.20)	46.67	92.70
'ML' Raking	Std. Lin. (1)	390.64	-121.98 (0.91)	125.31	86.80
	Std. Lin. (2)	391.10	-121.53 (0.91)	124.88	86.80
	Std. Lin. (3)	392.72	-119.91 (0.91)	123.32	87.10
	Jackk. Lin. (1)	486.91	-25.72 (1.24)	46.96	92.50
	Jackk. Lin. (2)	484.11	-28.52 (1.21)	47.72	92.50
	Jackk. Lin. (3)	482.83	-29.79 (1.19)	48.12	92.50

(1) Design Weights ω_i to compute Beta
(2) Calibration Weights w_i to compute Beta
(3) Alternative Calibration Weights $w\ _i^*$ to compute Beta

Finally, the table below shows relative bias of the standard error of the three point estimators under study, total number of unemployed people, employed and inactive.

Table 6.3 shows here that the true standard error is always underestimated, but never by more than 5.81% when the jackknife linearization method is used. In the case of the standard method, there is an important increase in the percent relative bias when considering non-response models, with values ranging from 16.22% to 23.80%. Even for the case of complete response, a decrease in the percent relative bias is observed when employing the jackknife linearization method rather than the standard method.

Once again, little change is observed in the percent relative bias of the standard error estimators when calibrated instead of initial weights are used to compute the estimated regression coefficients. No clear pattern is observed.

Table 6.3: Relative Bias of Standard Error Estimators of Unemployed, Employed and Inactive Totals, R = 1,000, data from LFS.

Estimator	Variance Estimation Method	Relative Bias of Standard Error		
		Unemployed	Employed	Inactive
Complete Response:				
GREG	Std. Linearization (1)	-6.54	-2.11	1.31
	Std. Linearization (2)	-6.47	-2.05	1.39
	Jackk. Linearization (1)	-4.90	-1.01	2.64
	Jackk. Linearization (2)	-5.08	-1.15	2.40
Classical Raking	Std. Linearization (1)	-6.66	-2.09	1.39
	Std. Linearization (2)	-6.60	-2.03	1.47
	Jackk. Linearization (1)	-5.00	-0.97	2.73
	Jackk. Linearization (2)	-5.21	-1.12	2.48
'ML' Raking	Std. Linearization (1)	-6.87	-2.06	1.45
	Std. Linearization (2)	-6.80	-1.99	1.53
	Std. Linearization (3)	-6.57	-1.83	1.80
	Jackk. Linearization (1)	-5.10	-0.86	2.89
	Jackk. Linearization (2)	-5.38	-1.07	2.57
	Jackk. Linearization (3)	-5.52	-1.17	2.44
Multiplicative Non-response:				
GREG	Std. Linearization (1)	-23.07	-20.10	-16.81
	Std. Linearization (2)	–	–	–
	Jackk. Linearization (1)	-4.84	-2.93	1.05
	Jackk. Linearization (2)	–	–	–
Classical Raking	Std. Linearization (1)	-23.17	-20.07	-16.71
	Std. Linearization (2)	-23.08	-19.99	-16.60
	Jackk. Linearization (1)	-4.90	-2.86	1.21
	Jackk. Linearization (2)	-5.27	-3.11	0.84
'ML' Raking	Std. Linearization (1)	-23.35	-20.09	-16.68
	Std. Linearization (2)	-23.26	-20.01	-16.57
	Std. Linearization (3)	-22.94	-19.79	-16.22
	Jackk. Linearization (1)	-4.86	-2.75	1.39
	Jackk. Linearization (2)	-5.43	-3.12	0.90
	Jackk. Linearization (3)	-5.69	-3.31	0.70

Table 6.3 continued on next page

6.2 The German Sample Survey of Income and Expenditure

6.2.1 Properties of the point estimators for alternative calibration methods under different non-response models

Biases, standard errors, root mean square roots, number of negative weights and very large weights for the studied estimators of the total income in our artificial population,

Continuation of Table 6.3

Estimator	Variance Estimation Method	Relative Bias of Standard Error		
		Unemployed	Employed	Inactive
Additive Non-response:				
GREG	Std. Linearization (1)	-23.44	-21.19	-17.86
	Std. Linearization (2)	–	–	–
	Jackk. Linearization (1)	-4.87	-3.87	0.16
	Jackk. Linearization (2)	–	–	–
Classical Raking	Std. Linearization (1)	-23.57	-21.14	-17.72
	Std. Linearization (2)	-23.49	-21.07	-17.62
	Jackk. Linearization (1)	-4.98	-3.79	0.37
	Jackk. Linearization (2)	-5.34	-4.04	0.01
'ML' Raking	Std. Linearization (1)	-23.80	-21.15	-17.62
	Std. Linearization (2)	-23.71	-21.07	-17.52
	Std. Linearization (3)	-23.39	-20.86	-17.17
	Jackk. Linearization (1)	-5.02	-3.66	0.63
	Jackk. Linearization (2)	-5.56	-4.02	0.15
	Jackk. Linearization (3)	-5.81	-4.20	-0.04

(1) Design Weights ω_i to compute Beta
(2) Calibration Weights w_i to compute Beta
(3) Alternative Calibration Weights w_i* to compute Beta

using the SIE data described, and considering the two non-response models presented in Chapter 5, Subsection 5.2.1, as well as complete response, are given in Table 6.4. Numbers are rounded to the nearest two decimals.

6.2.2 Properties of the variance estimators for alternative calibration methods under different non-response models

The average value of each standard error estimator $\widehat{S}\text{E} = \sqrt{\widehat{V}}$ for the alternatives total income estimators, under different non-response models, is shown in the Table 6.5. This table also summarizes the estimation bias for each standard error estimator, the root mean square error of these standard errors and the coverage confidence interval percentage.

Table 6.4: Simulation Bias of Income Total Estimate, Standard Error, Root Mean Square Error, number of negatives and very large calibration weights, R = 1,000, data from SIE.

Non-response Model/ Point Estimator	Bias (standard error)	"True" Standard Error	Root Mean Square Error	Number of Negative Weights	Number of Very Large Weights
Complete Response:					
GREG	196,243 (325,908)	10,306,105	10,307,973	0	0
Classical Raking	202,912 (326,054)	10,310,728	10,312,724	0	0
'ML' Raking	209,854 (326,187)	10,314,940	10,317,074	0	0
Multiplicative non-response:					
GREG	403,140 (415,304)	13,133,059	13,139,245	0	0
Classical Raking	398,971 (415,302)	13,133,010	13,139,069	0	0
'ML' Raking	423,073 (415,250)	13,131,346	13,138,160	0	1
Additive non-response:					
GREG	409,628 (416,442)	13,169,055	13,175,424	0	0
Classical Raking	407,664 (416,405)	13,167,872	13,174,181	0	0
'ML' Raking	434,233 (416,321)	13,165,234	13,172,393	0	1

Table 6.5: Expected Income Standard Error Estimate, Bias, Root Mean Square Error and Coverage Confidence Interval, R = 1,000, data from SIE.

Estimator	Variance Estimation Method	Expected Standard Error Estimate	Bias of SE Estimate (s.e.)	RMSE of SE Estimate	Coverage Confidence Interval (%)
Complete Response:					
GREG	Standard (1)	10,345,297	39,193 (6,920)	222,310	95.20
	Standard (2)	10,345,643	39,539 (6,921)	222,391	95.20
	Jackknife (1)	10,377,089	70,984 (7,097)	235,373	95.30
	Jackknife (2)	10,376,037	69,933 (7,093)	234,943	95.30
Classical Raking	Standard (1)	10,345,297	34,570 (6,920)	221,542	95.20
	Standard (2)	10,345,633	34,905 (6,920)	221,610	95.20
	Jackknife (1)	10,377,233	66,505 (7,103)	234,260	95.30
	Jackknife (2)	10,376,162	65,435 (7,100)	233,851	95.30
‘ML’ Raking	Standard (1)	10,345,297	30,357 (6,920)	220,924	95.20
	Standard (2)	10,345,635	30,695 (6,920)	220,985	95.20
	Standard (3)	10,346,710	31,771 (6,922)	221,191	95.20
	Jackknife (1)	10,377,624	62,684 (7,111)	233,429	95.30
	Jackknife (2)	10,376,475	61,535 (7,107)	233,013	95.30
	Jackknife (3)	10,376,062	61,122 (7,106)	232,869	95.30
Multiplicative non-response:					
GREG	Standard (1)	8,138,286	-4,994,773 (7,553)	5,000,480	77.90
	Standard (2)	8,139,051	-4,994,009 (7,553)	4,999,717	77.90
	Jackknife (1)	13,193,566	60,507 (12,726)	406,968	95.00
	Jackknife (2)	13,190,040	56,981 (12,704)	405,762	95.00
Classical Raking	Standard (1)	8,138,286	-4,994,723 (7,553)	5,000,431	77.50
	Standard (2)	8,139,029	-4,993,981 (7,553)	4,999,689	77.50
	Jackknife (1)	13,192,989	59,979 (12,755)	407,793	94.90
	Jackknife (2)	13,189,211	56,201 (12,732)	406,514	94.90
‘ML’ Raking	Standard (1)	8,138,286	-4,993,059 (7,553)	4,998,769	77.50
	Standard (2)	8,139,063	-4,992,283 (7,553)	4,997,993	77.50
	Standard (3)	8,141,784	-4,989,561 (7,557)	4,995,281	77.50
	Jackknife (1)	13,194,137	62,792 (12,832)	410,618	94.90
	Jackknife (2)	13,188,982	57,636 (12,783)	408,312	94.90
	Jackknife (3)	13,186,969	55,623 (12,769)	407,606	94.90

Table 6.5 continued on next page

Continuation of Table 6.5

Estimator	Variance Estimation Method	Expected Standard Error Estimate	Bias of SE Estimate (s.e.)	RMSE of SE Estimate	Coverage Confidence Interval (%)
Additive non-response:					
GREG	Standard (1)	8,118,509	-5,050,546 (7,581)	5,056,233	77.80
	Standard (2)	8,119,297	-5,049,757 (7,581)	5,055,445	77.80
	Jackknife (1)	13,221,553	52,498 (12,774)	407,334	94.60
	Jackknife (2)	13,217,936	48,881 (12,750)	406,138	94.60
Classical Raking	Standard (1)	8,118,509	-5,049,363 (7,581)	5,055,051	77.80
	Standard (2)	8,119,273	-5,048,598 (7,581)	5,054,287	77.80
	Jackknife (1)	13,221,205	53,334 (12,802)	408,322	94.70
	Jackknife (2)	13,217,331	49,459 (12,777)	407,048	94.70
‘ML’ Raking	Standard (1)	8,118,509	-5,046,725 (7,581)	5,052,416	77.90
	Standard (2)	8,119,307	-5,045,927 (7,581)	5,051,618	77.90
	Standard (3)	8,122,042	-5,043,191 (7,583)	5,048,890	78.00
	Jackknife (1)	13,222,610	57,376 (12,876)	411,190	94.80
	Jackknife (2)	13,217,337	52,104 (12,825)	408,891	94.80
	Jackknife (3)	13,215,328	50,095 (12,812)	408,223	94.80

(1) Design Weights ω_i to compute Beta
(2) Calibration Weights w_i to compute Beta
(3) Alternative Calibration Weights w_i* to compute Beta

Table 6.6: Relative Bias of Variance Estimators of Expenditure and Income Totals, R = 1,000, data from SIE.

Estimator	Variance Estimation Method	Relative Bias	
		Expenditure	Income
Complete Response:			
GREG	Standard (1)	-1.57	0.38
	Standard (2)	-1.57	0.38
	Jackknife (1)	-1.12	0.69
	Jackknife (2)	-1.13	0.69
Classical Raking	Standard (1)	-1.58	0.34
	Standard (2)	-1.57	0.34
	Jackknife (1)	-1.12	0.65
	Jackknife (2)	-1.13	0.63
'ML' Raking	Standard (1)	-1.58	0.29
	Standard (2)	-1.58	0.30
	Standard (3)	-1.57	0.31
	Jackknife (1)	-1.11	0.61
	Jackknife (2)	-1.13	0.60
	Jackknife (3)	-1.13	0.59
Multiplicative Non-response:			
GREG	Standard (1)	-37.40	-38.03
	Standard (2)	-37.39	-38.03
	Jackknife (1)	0.34	0.46
	Jackknife (2)	0.31	0.43
Classical Raking	Standard (1)	-37.42	-38.03
	Standard (2)	-37.42	-38.03
	Jackknife (1)	0.31	0.46
	Jackknife (2)	0.28	0.43
'ML' Raking	Standard (1)	-37.50	-38.02
	Standard (2)	-37.49	-38.02
	Standard (3)	-37.47	-38.00
	Jackknife (1)	0.23	0.48
	Jackknife (2)	0.17	0.44
	Jackknife (3)	0.15	0.42

Table 6.6 continued on next page

Continuation of Table 6.6

Estimator	Variance Estimation Method	Relative Bias	
		Expenditure	Income
Additive Non-response:			
GREG	Standard (1)	-37.49	-38.35
	Standard (2)	-37.49	-38.35
	Jackknife (1)	0.72	0.40
	Jackknife (2)	0.69	0.37
Classical Raking	Standard (1)	-37.52	-38.35
	Standard (2)	-37.51	-38.34
	Jackknife (1)	0.69	0.41
	Jackknife (2)	0.66	0.38
'ML' Raking	Standard (1)	-37.59	-38.33
	Standard (2)	-37.59	-38.33
	Standard (3)	-37.56	-38.31
	Jackknife (1)	0.61	0.44
	Jackknife (2)	0.55	0.40
	Jackknife (3)	0.53	0.38

(1) Design Weights ω_i to compute Beta
(2) Calibration Weights w_i to compute Beta
(3) Alternative Calibration Weights w_i * to compute Beta

Chapter 7

Conclusions

The simulation study showed little difference between the bias or variance properties of the three calibration estimators considered: the GREG estimator, the classical raking estimator and the maximum likelihood raking estimator. Some small differences in the distribution of extreme weights were observed. A few negative weights were observed for the GREG estimator, whereas weights were necessarily positive for both raking estimators. Some very large weights were observed for the maximum likelihood raking estimator, suggesting no advantage in this method, despite the fact that one might have expected it to demonstrate different bias properties in the presence of non-response.

Amongst the variance estimators, the main finding was the contrast between the 'standard' linearization variance estimator which weights residuals by the design weight and the 'jackknife linearization' variance estimator which weights residuals by the calibrated weight. It was found that the latter variance estimator tended always to have reduced bias and that this effect was very marked in the presence of non-response, when the former estimator could be severely biased. The bias of the jackknife linearization variance estimator was generally small and the coverage level of the associated confidence intervals was generally close to the nominal coverage.

Alternative ways of weighting the regression coefficients when calculating the residuals in the linearization variance estimator were considered but little effect was observed and there was no evidence that this choice is important.

In general, the findings for the categorical variables in the British Labour Force Survey were remarkably similar to the findings for the continuous variables in the German Income and Expenditure survey.

Appendix A

Distribution of population and sample sizes of the LFS across strata

Table A.1: Distribution of population and sample sizes across strata

Strata	N_h	n_h	f
1	1216	25	0.0206
2	2139	45	0.0210
3	1387	29	0.0209
4	2143	45	0.0210
5	1689	35	0.0207
6	4263	89	0.0209
7	2271	47	0.0207
8	2230	46	0.0206
9	4041	84	0.0208
10	11127	232	0.0209
11	4959	103	0.0208
12	2719	57	0.0210
13	2840	59	0.0208
14	2513	52	0.0207
15	1374	29	0.0211
16	2430	51	0.0210
17	2975	62	0.0208
18	2551	53	0.0208
19	3247	68	0.0209
Total	58114	1211	0.0208

Appendix B

S-Plus® Functions

```
#FUNCTION CALIBRATION
#==================
# This function computes generalised linear regression weights
# (Function 1), raking weights # (Function 2) and 'maximum
# likelihood' raking weights (Function 4).
#
# INPUT:
#        Distance.Function.Number -- (1) GREG, (2) Classical Raking,
#                                    (4) ML raking;
#        F.func - function defining calibration distance from
#                 DEVILLE and SÄRNDAL (1992);
#        Flinha.func - gradient of function defining calibration distance;
#        Mat.X.s - sample data matrix for auxiliary variables to be used
#                  for calibration;
#        Pop.Total.X - vector of population totals for calibration;
#        Vect.Pi.s - vector of inclusion probabilities for each
#                    sample unit;
#        limit - maximum number of iterations to perform if convergence
#                not achieved.
#
# OUTPUT:
#        The vector of calibrated weights;
#

''CALIBRATION'' <- function(Distance.Function.Number, Mat.X.s,
                            Pop.Total.X, Vect.Pi.s) {

    Sample.Size <- length(Vect.Pi.s)
    Design.Weights <- 1/Vect.Pi.s
    f <- rep(1, times = Sample.Size)
    Mat.X.s <- as.matrix(Mat.X.s)
    Pop.Total.X <- as.vector(Pop.Total.X)

    if(Distance.Function.Number == 1) {
        Calibration.Weights.s <- CalibDeville.f(F1.f, F1linha.f,
        Mat.X.s, Pop.Total.X, Design.Weights, f)
    }
```

```
    if(Distance.Function.Number == 2) {
        Calibration.Weights.s <- CalibDeville.f(F2.f, F2linha.f,
        Mat.X.s, Pop.Total.X, Design.Weights, f)
    }

    if(Distance.Function.Number == 4) {
        Calibration.Weights.s <- CalibDeville.4(F4.f, F4linha.f,
        Mat.X.s, Pop.Total.X, Design.Weights, f)
    }

# OUTPUT
    Calibration.Weights.s

 }



''CalibDeville.f'' <- function(F.func, Flinha.func, X, Tx, d, f, limit,
                                eps = 1e-005, ets = 1e-005) {

# Residual function that defines nonlinear system to be solved to
# obtain lambda:
    residuos <- function(lamb, F.func, X, Tx, d, f) {


# Function for computing residuals of linear model given X

# Converts data into proper object classes

    lamb <- matrix(lamb, ncol = 1)
    d <- matrix(d, ncol = 1)
    f <- matrix(f, ncol = 1)
    Tx <- matrix(Tx, ncol = 1)        # Compute required residuals
    u <- X  %*%  lamb


# Compute estimated total of x variables using HT estimator
    Txpi <- t(X)  %*%  d
    t(X)  %*%  ((F.func(u * f) - 1) * d) - (Tx - Txpi)
 }


# Function to compute Jacobian of specified distance function needed
# for solving for lambda
    jacobiano <- function(lamb, Flinha.func, X, Tx, d, f) {


# Converts data into proper object classes
    lamb <- matrix(lamb, ncol = 1)
    d <- matrix(d, ncol = 1)
    f <- matrix(f, ncol = 1)
    Tx <- matrix(Tx, ncol = 1)
    u <- X  %*%  lamb
    t(X) %*%  diag(as.vector(d * (Flinha.func(f * u) * f)),
                     nrow = length(d)) %*%  X
 }
```

```
# Compute estimated total of x variables using HT estimator
    Txpi <- t(X)  %*%  d           # Initializing lambda
    lamb <- GINVERSE(t(X)  %*%  diag(d)  %*%  X)  %*%  (Tx - Txpi)


# Initializing values required for solution
    Func <- residuos(lamb, F.func, X, Tx, d, f)
    Jota <- jacobiano(lamb, Flinha.func, X, Tx, d, f)
    delta <- GINVERSE(Jota)  %*%  ( - Func)
    it <- 1 {\#} Computes lambda by Newton's method

    while(((sum(delta\^{}2) >= eps $\vert \vert $ sum(abs(Func)) >= ets)
      && (it <- it + 1) < limit)) {

        Func <- residuos(lamb, F.func, X, Tx, d, f)
        Jota <- jacobiano(lamb, Flinha.func, X, Tx, d, f)
        delta <- GINVERSE(Jota)  %*%  ( - Func)
        lamb <- lamb + delta
 }

    u <- X  %*%  lamb {\#} Calibration weights
    w.v <- as.vector(d * F.func(f * u))


# Defines output to be provided by function
    return(w.v)

 }



''CalibDeville.4''<- function(F.func, Flinha.func, X, Tx, d, f, limit,
                              eps = 1e-005, ets = 1e-005) {

# Residual function that defines nonlinear system to be solved to
# obtain lambda:
    residuos <- function(lamb, F.func, X, Tx, d, f) {


# Function for computing residuals of linear model given X

# Converts data into proper object classes
    lamb <- matrix(lamb, ncol = 1)
    d <- matrix(d, ncol = 1)
    f <- matrix(f, ncol = 1)
    Tx <- matrix(Tx, ncol = 1)       # Compute required residuals
    u <- X  %*%  lamb


# Compute estimated total of x variables using HT estimator
    Txpi <- t(X)  %*%  d
    t(X)  %*%  ((F.func(u * f) - 1) * d) - (Tx - Txpi)

 }
```

```
# Function to compute Jacobian of specified distance function needed for
# solving for lambda

    jacobiano <- function(lamb, Flinha.func, X, Tx, d, f) {


# Converts data into proper object classes
    lamb <- matrix(lamb, ncol = 1)
    d <- matrix(d, ncol = 1)
    f <- matrix(f, ncol = 1)
    Tx <- matrix(Tx, ncol = 1)
    u <- X %*% lamb
    t(X) %*% diag(as.vector(d * (Flinha.func(f * u) * f)), nrow =
    length(d)) %*% X

 }


# Compute estimated total of x variables using HT estimator
    Txpi <- t(X) %*% d {\#} Initializing lambda
    lamb <- GINVERSE(t(X) %*% diag(d) %*% X) %*% (Tx - Txpi)
    u <- X %*% lamb {\#} Calibration weights
    h <- max(u)
    if (h > 0.99) {
        position <- cbind(u,c(1:length(u)))
        x.value <- X[position[u==h,2],]
        tita <- 0.99/(x.value %*% lamb)
        lamb <- c(tita) * lamb
    }


# Initializing values required for solution
    Func <- residuos(lamb, F.func, X, Tx, d, f)
    Jota <- jacobiano(lamb, Flinha.func, X, Tx, d, f)
    delta <- GINVERSE(Jota) %*% ( - Func)
    it <- 1 {\#} Computes lambda by Newton's method

    while(((sum(delta\^{}2) >= eps $\vert \vert $ sum(abs(Func)) >= ets)
                                && (it <- it + 1) < limit)) {
        Func <- residuos(lamb, F.func, X, Tx, d, f)
        Jota <- jacobiano(lamb, Flinha.func, X, Tx, d, f)
        delta <- GINVERSE(Jota) %*% ( - Func)
        lamb <- lamb + delta
        u <- X %*% lamb # Calibration weights

        h <- max(u)

# Check X %*% lamb belongs to I $\in $ R
    if (h > 0.99) {
        lamb.before <- lamb - delta
        position <- cbind(u,c(1:length(u)))
        x.value <- X[position[u==h,2],]
        tita <- (0.99 - x.value %*% lamb.before)/(x.value %*
                 % (lamb - lamb.before))
        lamb <- lamb.before + c(tita) * (lamb - lamb.before)
    }
}
```

```
    u <- X  %*%  lamb {\#} Calibration weights
    w.v <- as.vector(d * F.func(f * u))


# Defines output to be provided by function
    return(w.v)

}

    ''F1.f''<- function(u) {
        1 + u
    }

    ''F2.f''<- function(u) {
        exp(u)
    }

    ''F4.f''<- function(u) {
        (1 - u)\^{}(-1)
    }

    ''F1linha.f''<- function(u) {
        rep(1, length(u))
    }

    ''F2linha.f''<- function(u) {
        exp(u)
    }

    ''F4linha.f''<- function(u) {
        (1 - u)\^{}(-2)
    }

    ''GINVERSE''<- function(x, tol = sqrt(.Machine$double.eps)) {
        if(length(dim(x)) > 2)
            stop(''x must be a matrix or vector'')
            svdX <- svd(x)
        if(is.complex(x))
            svdX$u <- Conj(svdX$u)
            NotZero <- svdX$d > tol * svdX$d[1]
            ans <- if(all(NotZero)) svdX$v  %*%  ((1/svdX$d) * t(svdX$u)
                )
                    else if(!any(NotZero)) {
                if(is.matrix(x))
                array(0, dim(x)[2:1])
                else matrix(0, 1, length(x))
    }

    else svdX$v[, NotZero]  %*%  ((1/svdX$d[NotZero]) * t(svdX$u[,
        NotZero]))
    attr(ans, ''rank'') <- sum(NotZero)
    ans
 }
```

References

Binder, D. A. and Théberge, A. (1988): Estimating the variance of raking-ratio estimators. *Canadian Journal of Statistics* **16**, 47–55.

Brackstone, G. J. and Rao, J. N. K. (1979): An investigation of raking ratio estimators. *Sankyā, Series C* **41**, 97–114.

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

Demnati, A. and Rao, J. N. K. (2004): Linearization variance estimators for survey data. to appear in: Survey Methodology.

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

Deville, J. C., Särndal, C. E. and Sautory, O. (1993): Generalized raking procedures in survey sampling. *Journal of the American Statistical Association* **88**, 1013–1020.

Foster, K. (1998): Evaluating non-response on household-surveys. Technical report, GSS Methodology Series, Office for National Statistics, London.

Fuller, W. A. (2002): Regression estimating for survey samples. *Survey Methodology* **28**, 5–23.

Ireland, C. T. and Kullback, S. (1968): Contingency tables with given marginals. *Biometrika* **55**, 179–188.

Kalton, G. and Flores-Cervantes, I. (2003): Weighting methods. *Journal of Official Statistics* **19**, 81–98.

Office for National Statistics (1998): Labour Force Survey user guide, volume 1: Background and methodology. Technical report, Office for National Statistics, London.

Rao, J. N. K. and Tausi, M. (2003): Estimation function jackknife variance estimators under stratified multistage sampling. to appear in: Communications in Statistics.

Rust, K. F. and Rao, J. N. K. (1996): Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research* **5**, 283–310.

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